Financial Engineering Advanced Background Series

Published or forthcoming

1. A Primer for the Mathematics of Financial Engineering, by Dan Stefanica
2. Numerical Linear Algebra Methods for Financial Engineering Applications, by Dan Stefanica
3. A Probability Primer for Mathematical Finance, by Elena Kosygina
4. Differential Equations with Numerical Methods for Financial Engineering, by Dan Stefanica

A PRIMER for the MATHEMATICS of FINANCIAL ENGINEERING

DAN STEFANICA

Baruch College
City University of New York

FE PRESS
New York

FE PRESS
New York

www.fepress.org
Information on this title: www.fepress.org/mathematical_primer

First published 2008

Printed in the United States of America

ISBN-13 978-0-9797576-0-0
ISBN-10 0-9797576-0-6

To Miriam

and

to Rianna

Contents

List of Tables

Preface

The use of quantitative models in trading has grown tremendously in recent years, and seems likely to grow at similar speeds in the future, due to the availability of ever faster and cheaper computing power. Although many books are available for anyone interested in learning about the mathematical models used in the financial industry, most of these books target either the finance practitioner, and are lighter on rigorous mathematical fundamentals, or the academic scientist, and use high-level mathematics without a clear presentation of its direct financial applications.

This book is meant to build the solid mathematical foundation required to understand these quantitative models, while presenting a large number of financial applications. Examples range from Put-Call parity, bond duration and convexity, and the Black–Scholes model, to more advanced topics, such as the numerical estimation of the Greeks, implied volatility, and bootstrapping for finding interest rate curves. On the mathematical side, useful but sometimes overlooked topics are presented in detail: differentiating integrals with respect to nonconstant integral limits, numerical approximation of definite integrals, convergence of Taylor series, finite difference approximations, Stirling's formula, Lagrange multipliers, polar coordinates, and Newton's method for multidimensional problems. The book was designed so that someone with a solid knowledge of Calculus should be able to understand all the topics presented.

Every chapter concludes with exercises that are a mix of mathematical and financial questions, with comments regarding their relevance to practice and to more advanced topics. Many of these exercises are, in fact, questions that are frequently asked in interviews for quantitative jobs in financial institutions, and some are constructed in a sequential fashion, building upon each other, as is often the case at interviews. Complete solutions to most of the exercises can be found at http://www.fepress.org/

This book can be used as a companion to any more advanced quantitative finance book. It also makes a good reference book for mathematical topics that are frequently assumed to be known in other texts, such as Taylor expansions, Lagrange multipliers, finite difference approximations, and numerical methods for solving nonlinear equations.

This book should be useful to a large audience:

- Prospective students for financial engineering (or mathematical finance)

programs will find that the knowledge contained in this book is fundamental for their understanding of more advanced courses on numerical methods for finance and stochastic calculus, while some of the exercises will give them a flavor of what interviewing for jobs upon graduation might be like.

• For finance practitioners, while parts of the book will be light reading, the book will also provide new mathematical connections (or present them in a new light) between financial instruments and models used in practice, and will do so in a rigorous and concise manner.

• For academics teaching financial mathematics courses, and for their students, this is a rigorous reference book for the mathematical topics required in these courses.

• For professionals interested in a career in finance with emphasis on quantitative skills, the book can be used as a stepping stone toward that goal, by building a solid mathematical foundation for further studies, as well as providing a first insight in the world of quantitative finance.

The material in this book has been used for a mathematics refresher course for students entering the Financial Engineering Masters Program (MFE) at Baruch College, City University of New York. Studying this material before entering the program provided the students with a solid background and played an important role in making them successful graduates: over 90 percent of the graduates of the Baruch MFE Program are currently employed in the financial industry.

The author has been the Director of the Baruch College MFE Program[1] since its inception in 2002. This position gave him the privilege to interact with generations of students, who were exceptional not only in terms of knowledge and ability, but foremost as very special friends and colleagues. The connection built during their studies has continued over the years, and as alumni of the program their contribution to the continued success of our students has been tremendous.

This is the first in a series of books containing mathematical background needed for financial engineering applications, to be followed by books in Numerical Linear Algebra, Probability, and Differential Equations.

Dan Stefanica

New York, 2008

[1]Baruch MFE Program web page: http://www.baruch.cuny.edu/math/masters.html
QuantNetwork student forum web page: http://www.quantnet.org/forum/index.php

Acknowledgments

I have spent several wonderful years at Baruch College, as Director of the Financial Engineering Masters Program. Working with so many talented students was a privilege, as well as a learning experience in itself, and seeing a strong community develop around the MFE program was incredibly rewarding. This book is by all accounts a direct result of interacting with our students and alumni, and I am truly grateful to all of them for this.

The strong commitment of the administration of Baruch College to support the MFE program and provide the best educational environment to our students was essential to all aspects of our success, and permeated to creating the opportunity for this book to be written.

I learned a lot from working alongside my colleagues in the mathematics department and from many conversations with practitioners from the financial industry. Special thanks are due to Elena Kosygina and Sherman Wong, as well as to my good friends Peter Carr and Salih Neftci. The title of the book was suggested by Emanuel Derman, and is more euphonious than any previously considered alternatives.

Many students have looked over ever–changing versions of the book, and their help and encouragement were greatly appreciated. The knowledgeable comments and suggestions of Robert Spruill are reflected in the final version of the book, as are exercises suggested by Sudhanshu Pardasani. Andy Nguyen continued his tremendous support both on QuantNet.org, hosting the problems solutions, and on the fepress.org website. The art for the book cover is due to Max Rumyantsev. The final effort of proofreading the material was spareheaded by Vadim Nagaev, Muting Ren, Rachit Gupta, Claudia Li, Sunny Lu, Andrey Shvets, Vic Siqiao, and Frank Zheng.

I would have never gotten past the lecture notes stage without tremendous support and understanding from my family. Their smiling presence and unwavering support brightened up my efforts and made them worthwhile.

This book is dedicated to the two ladies in my life.

Dan Stefanica

New York, 2008

How to Use This Book

While we expect a large audience to find this book useful, the approach to reading the book will be different depending on the background and goals of the reader.

Prospective students for financial engineering or mathematical finance programs should find the study of this book very rewarding, as it will give them a head start in their studies, and will provide a reference book throughout their course of study. Building a solid base for further study is of tremendous importance. This book teaches core concepts important for a successful learning experience in financial engineering graduate programs.

Instructors of quantitative finance courses will find the mathematical topics and their treatment to be of greatest value, and could use the book as a reference text for a more advanced treatment of the mathematical content of the course they are teaching.

Instructors of financial mathematics courses will find that the exercises in the book provide novel assignment ideas. Also, some topics might be non-traditional for such courses, and could be useful to include or mention in the course.

Finance practitioners should enjoy the rigor of the mathematical presentation, while finding the financial examples interesting, and the exercises a potential source for interview questions.

The book was written with the aim of ensuring that anyone thoroughly studying it will have a strong base for further study and full understanding of the mathematical models used in finance.

A point of caution: there is a significant difference between studying a book and merely reading it. To benefit fully from this book, all exercises should be attempted, and the material should be learned as if for an exam. Many of the exercises have particular relevance for people who will interview for quantitative jobs, as they have a flavor similar to questions that are currently asked at such interviews.

The book is sequential in its presentation, with the exception of Chapter 0, which can be skipped over and used as a collection of reference topics.

Chapter 0

Mathematical preliminaries

Even and odd functions.

Useful sums with interesting proofs.

Sequences satisfying linear recursions.

The "Big O" and "little o" notations.

This chapter is a collection of topics that are needed later on in the book, and may be skipped over in a first reading. It is also the only chapter of the book where no financial applications are presented.

Nonetheless, some of the topics in this chapter are rather subtle from a mathematical standpoint, and understanding their treatment is instructive. In particular, we include a discussion of the "Big O" and "little o" notations, i.e., $O(\cdot)$ and $o(\cdot)$, which are often a source of confusion.

0.1 Even and odd functions

Even and odd functions are special families of functions whose graphs exhibit special symmetries. We present several simple properties of these functions which will be used subsequently.

Definition 0.1. *The function $f : \mathbb{R} \to \mathbb{R}$ is an even function if and only if*

$$f(-x) \;=\; f(x), \quad \forall\, x \in \mathbb{R}. \tag{1}$$

The graph of any even function is symmetric with respect to the y–axis.

Example: The density function $f(x)$ of the standard normal variable, i.e.,

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}},$$

is an even function, since

$$f(-x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-x)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = f(x);$$

see section 3.3 for more properties of this function. □

Lemma 0.1. *Let $f(x)$ be an integrable even function. Then,*

$$\int_{-a}^{0} f(x)\,dx = \int_{0}^{a} f(x)\,dx, \quad \forall\, a \in \mathbb{R}, \tag{2}$$

and therefore

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x), \quad \forall\, a \in \mathbb{R}. \tag{3}$$

Moreover, if $\int_0^\infty f(x)\,dx$ exists, then

$$\int_{-\infty}^{0} f(x)\,dx = \int_{0}^{\infty} f(x)\,dx, \tag{4}$$

and

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\int_{0}^{\infty} f(x). \tag{5}$$

Proof. Use the substitution $x = -y$ for the integral on the left hand side of (2). The end points $x = -a$ and $x = 0$ change into $y = a$ and $y = 0$, respectively, and $dx = -dy$. We conclude that

$$\int_{-a}^{0} f(x)\,dx = \int_{a}^{0} f(-y)\,(-dy) = \int_{0}^{a} f(-y)\,dy = \int_{0}^{a} f(y)\,dy, \tag{6}$$

since $f(-y) = f(y)$; cf. (1). Note that y is just an integrating variable. Therefore, we can replace y by x in (6) to obtain

$$\int_{-a}^{0} f(x)\,dx = \int_{0}^{a} f(x)\,dx.$$

Then,

$$\int_{-a}^{a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx = 2\int_{0}^{a} f(x).$$

The results (4) and (5) follow from (2) and (3) using the definitions (2.5), (2.6), and (2.7) of improper integrals.

For example, the proof of (4) can be obtained using (2) as follows:

$$\begin{aligned}\int_{-\infty}^{0} f(x)\,dx &= \lim_{t\to-\infty}\int_{t}^{0} f(x)\,dx = \lim_{t\to-\infty}\int_{0}^{-t} f(x)\,dx \\ &= \lim_{t\to\infty}\int_{0}^{t} f(x)\,dx = \int_{0}^{\infty} f(x)\,dx.\end{aligned}$$

□

Definition 0.2. *The function $f : \mathbb{R} \to \mathbb{R}$ is an odd function if and only if*

$$f(-x) = -f(x), \quad \forall\, x \in \mathbb{R}. \tag{7}$$

If we let $x = 0$ in (7), we find that $f(0) = 0$ for any odd function $f(x)$. Also, the graph of any odd function is symmetric with respect to the point $(0, 0)$.

Lemma 0.2. *Let $f(x)$ be an integrable odd function. Then,*

$$\int_{-a}^{a} f(x)\,dx = 0, \quad \forall\, a \in \mathbb{R}. \tag{8}$$

Moreover, if $\int_0^\infty f(x)\,dx$ exists, then

$$\int_{-\infty}^{\infty} f(x)\,dx = 0. \tag{9}$$

Proof. Use the substitution $x = -y$ for the integral from (8). The end points $x = -a$ and $x = a$ change into $y = a$ and $y = -a$, respectively, and $dx = -dy$. Therefore,

$$\int_{-a}^{a} f(x)\,dx = \int_{a}^{-a} f(-y)\,(-dy) = \int_{-a}^{a} f(-y)\,dy = -\int_{-a}^{a} f(y)\,dy, \tag{10}$$

since $f(-y) = -f(y)$; cf. (7). Since y is just an integrating variable, we can replace y by x in (10), and obtain that

$$\int_{-a}^{a} f(x)\,dx = -\int_{-a}^{a} f(x)\,dx.$$

We conclude that

$$\int_{-a}^{a} f(x)\,dx = 0.$$

The result of (9) follows from (8) and (2.10). □

0.2 Useful sums with interesting proofs

The following sums occur frequently in practice, e.g., when estimating the operation counts of numerical algorithms:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}; \tag{11}$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}; \tag{12}$$

$$\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2. \tag{13}$$

Using mathematical induction, it is easy to show that formulas (11–13) are correct. For example, for formula (13), a proof by induction can be given as follows: if $n = 1$, both sides of (13) are equal to 1. We assume that (13) holds for n and prove that (13) also holds for $n + 1$. In other words, we assume that

$$\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2, \tag{14}$$

and show that

$$\sum_{k=1}^{n+1} k^3 = \left(\frac{(n+1)(n+2)}{2}\right)^2. \tag{15}$$

From (14), and by a simple computation, we find that

$$\begin{aligned}\sum_{k=1}^{n+1} k^3 &= \sum_{k=1}^{n} k^3 + (n+1)^3 = \left(\frac{n(n+1)}{2}\right)^2 + (n+1)^3 \\ &= (n+1)^2\left(\frac{n^2}{4} + n + 1\right) = \frac{(n+1)^2(n^2+4n+4)}{4} \\ &= \left(\frac{(n+1)(n+2)}{2}\right)^2.\end{aligned}$$

In other words, (15) is proven, and therefore (13) is established for any $n \geq 1$, by induction.

While proving inequalities (11–13) by induction is not difficult, an interesting question is how are these formulas derived in the first place? In other words, how do we find out what the correct right hand sides of (11–13) are?

We present here two different methods for obtaining closed formulas for any sum of the form

$$S(n,i) = \sum_{k=1}^{n} k^i, \tag{16}$$

where $i \geq 1$ and $k \geq 1$ are positive integers.

First Method: Recall from the binomial formula that

$$(a+b)^m = \sum_{j=0}^{m} \binom{m}{j} a^j b^{m-j}, \tag{17}$$

for any real numbers a and b, and for any positive integer m. The term $\binom{m}{j}$ is the binomial coefficient defined as follows:

$$\binom{m}{j} = \frac{m!}{j!\,(m-j)!},$$

where the factorial of a positive integer k is defined as $k! = 1 \cdot 2 \cdot \ldots \cdot k$.

Using (17) for $a = k$, $b = 1$, and $m = i + 1$, where k and i are positive integers, we obtain that

$$(k+1)^{i+1} = \sum_{j=0}^{i+1} \binom{i+1}{j} k^j = k^{i+1} + \sum_{j=0}^{i} \binom{i+1}{j} k^j.$$

Therefore,

$$(k+1)^{i+1} - k^{i+1} = \sum_{j=0}^{i} \binom{i+1}{j} k^j. \tag{18}$$

Writing (18) for all positive integers $k = 1 : n$, summing over k, and using the notation from (16), we obtain that

$$\begin{aligned}(n+1)^{i+1} - 1 &= \sum_{k=1}^{n}\left(\sum_{j=0}^{i} \binom{i+1}{j} k^j\right) \\ &= \sum_{j=0}^{i} \binom{i+1}{j} \left(\sum_{k=1}^{n} k^j\right) \\ &= \sum_{j=0}^{i} \binom{i+1}{j} S(n,j) \\ &= (i+1)S(n,i) + \sum_{j=0}^{i-1} \binom{i+1}{j} S(n,j).\end{aligned}$$

We established the following recursive formula for $S(n,i) = \sum_{k=1}^{n} k^i$:

$$S(n,i) \;=\; \frac{1}{i+1}\left((n+1)^{i+1} - 1 \;-\; \sum_{j=0}^{i-1}\binom{i+1}{j}\, S(n,j)\right), \qquad (19)$$

for all $i \geq 1$. It is easy to see that, for $i = 0$,

$$S(n,0) \;=\; \sum_{k=1}^{n} k^0 \;=\; \sum_{k=1}^{n} 1 \;=\; n.$$

Example: Use the recursion formula (19) to compute $S(n,1) = \sum_{k=1}^{n} k$ and $S(n,2) = \sum_{k=1}^{n} k^2$.

Answer: Recall that $S(n,0) = n$. For $i = 1$, formula (19) becomes

$$\begin{aligned} S(n,1) &= \frac{1}{2}\left((n+1)^2 - 1 \;-\; \binom{2}{0}\, S(n,0)\right) \\ &= \frac{1}{2}\left((n+1)^2 - 1 - n\right) \;=\; \frac{n(n+1)}{2}, \end{aligned}$$

which is the same as formula (11).

For $i = 2$, formula (19) becomes

$$\begin{aligned} S(n,2) &= \frac{1}{3}\left((n+1)^3 - 1 \;-\; \sum_{j=0}^{1}\binom{3}{j}\, S(n,j)\right) \\ &= \frac{1}{3}\left(\,(n+1)^3 - 1 \;-\; S(n,0) \;-\; 3S(n,1)\,\right) \\ &= \frac{1}{3}\left((n+1)^3 - 1 \;-\; n \;-\; \frac{3n(n+1)}{2}\right) \\ &= \frac{1}{3}\cdot\frac{2n^3 + 3n^2 + n}{2} \;=\; \frac{n(n+1)(2n+1)}{6}, \end{aligned}$$

which is the same as formula (12). $\Box$

Second Method: Another method to compute $S(n,i) = \sum_{k=1}^{n} k^i$, for $i \geq 0$ positive integer is to find a closed formula for

$$T(n,j,x) \;=\; \sum_{k=1}^{n} k^j\, x^k,$$

where $j \geq 0$ is an integer and $x \in \mathbb{R}$, and then evaluate $T(n,j,x)$ at $x = 1$ to obtain

$$S(n,i) \;=\; T(n,i,1) \;=\; \sum_{k=1}^{n} k^i, \;\; \forall\, n \geq 1, \;\; i \geq 1. \qquad (20)$$

We provide a recursive formula for evaluating $T(n,j,x)$.

For $j = 0$, we find that $T(n,0,x) = \sum_{k=1}^{n} x^k = \sum_{k=0}^{n} x^k - 1$. Since

$$\sum_{k=0}^{n} x^k \;=\; \frac{1 - x^{n+1}}{1 - x}, \qquad (21)$$

which can be seen, e.g., by cross multiplication,

$$(1-x)\sum_{k=0}^{n} x^k \;=\; \sum_{k=0}^{n} (1-x)x^k \;=\; \sum_{k=0}^{n}\left(x^k - x^{k+1}\right) \;=\; 1 - x^{n+1},$$

it follows that

$$T(n,0,x) \;=\; \frac{1 - x^{n+1}}{1 - x} \;-\; 1, \;\; \forall\, n \geq 1, \;\; x \in \mathbb{R}. \qquad (22)$$

Note that

$$x\,\frac{d}{dx}(T(n,j,x)) \;=\; x\sum_{k=1}^{n} k^j \cdot kx^{k-1} \;=\; \sum_{k=1}^{n} k^{j+1}\, x^k \;=\; T(n,j+1,x).$$

Thus, the following recursion formula holds:

$$T(n,j+1,x) \;=\; x\,\frac{d}{dx}(T(n,j,x)), \;\; \forall\, j \geq 0. \qquad (23)$$

Formulas (22) and (23) can be used to evaluate $T(n,j,x)$ at any point $x \in \mathbb{R}$, and for any integers $n \geq 1$ and $j \geq 0$.

We note that evaluating $T(n,i,x)$ at $x = 1$, which is needed to compute $S(n,i)$, see (20), requires using l'Hôpital's rule to compute limits as $x \to 1$.

Example: Use the recursion formula (23) and the fact that $S(n,i) = T(n,i,1)$ for any positive integer i to compute $S(n,1) = \sum_{k=1}^{n} k$.

Answer: For $j = 0$, formula (23) becomes

$$T(n,1,x) \;=\; x\,\frac{d}{dx}(T(n,0,x)).$$

Using (22), it follows that

$$T(n,1,x) \;=\; \frac{x - (n+1)x^{n+1} + nx^{n+2}}{(1-x)^2}. \qquad (24)$$

Then, the value of $S(n,1) = T(n,1,1)$ can be obtained by computing the limit of the right hand side of (24) as $x \to 1$. Using l'Hôpital's rule, see

Theorem 1.8, we obtain that

$$
\begin{aligned}
S(n,1) &= T(n,1,1) = \lim_{x\to 1} \frac{x-(n+1)x^{n+1}+nx^{n+2}}{(1-x)^2} \\
&= \lim_{x\to 1} \frac{1-(n+1)^2x^n+n(n+2)x^{n+1}}{-2(1-x)} \\
&= \lim_{x\to 1} \frac{-n(n+1)^2x^{n-1}+n(n+1)(n+2)x^n}{2} \\
&= \frac{-n(n+1)^2+n(n+1)(n+2)}{2} = \frac{n(n+1)}{2},
\end{aligned}
$$

which is the same as formula (11). □

0.3 Sequences satisfying linear recursions

Definition 0.3. *A sequence $(x_n)_{n\geq 0}$ satisfies a linear recursion of order k if and only if there exist constants a_i, $i = 0 : k$, with $a_k \neq 0$, such that*

$$\sum_{i=0}^{k} a_i x_{n+i} = 0, \quad \forall\, n \geq 0. \tag{25}$$

The recursion (25) is called a linear recursion because of the following linearity properties:

(i) If the sequence $(x_n)_{n\geq 0}$ satisfies the linear recursion (25), then the sequence $(z_n)_{n\geq 0}$ given by

$$z_n = Cx_n, \quad \forall\, n \geq 0, \tag{26}$$

where C is an arbitrary constant, also satisfies the linear recursion (25).

(ii) If the sequences $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ satisfy the linear recursion (25), then the sequence $(z_n)_{n\geq 0}$ given by

$$z_n = x_n + y_n, \quad \forall\, n \geq 0, \tag{27}$$

also satisfies the linear recursion (25).

Note that, if the first k numbers of the sequence, i.e., $x_0, x_1, \ldots, x_{k-1}$, are specified, then all entries of the sequence are uniquely determined by the recursion formula (25): since $a_k \neq 0$, we can solve (25) for x_{n+k}, i.e.,

$$x_{n+k} = -\frac{1}{a_k} \sum_{i=0}^{k-1} a_i x_{n+i} \quad \forall\, n \geq 0. \tag{28}$$

If $x_0, x_1, \ldots, x_{k-1}$ are given, we find x_k by letting $n = 0$ in (28). Then $x_1, x_2, \ldots, x_k$ are known and we find x_{k+1} by letting $n = 1$ in (28), and so on.

In Theorem 0.1, we will present the general formula of x_n in terms of $x_0, x_1, \ldots, x_{k-1}$. To do so, we first define the characteristic polynomial[2] associated to a linear recursion.

Definition 0.4. *The characteristic polynomial $P(z)$ corresponding to the linear recursion $\sum_{i=0}^{k} a_i x_{n+i} = 0$, for all $n \geq 0$, is defined as*

$$P(z) = \sum_{i=0}^{k} a_i z^i. \tag{29}$$

Note that $P(z)$ is a polynomial of degree k, i.e., $\deg(P(z)) = k$. Recall that every polynomial of degree k with real coefficients has exactly k roots (which could be complex numbers), when counted with their multiplicities. More precisely, if $P(z)$ has p different roots λ_j, $j = 1 : p$, with $p \leq k$, and if $m(\lambda_j)$ denotes the multiplicity of the root λ_j, then $\sum_{j=1}^{p} m(\lambda_j) = k$.

Theorem 0.1. *Let $(x_n)_{n\geq 0}$ be a sequence satisfying the linear recursion*

$$\sum_{i=0}^{k} a_i x_{n+i} = 0, \quad \forall\, n \geq 0, \tag{30}$$

with $a_k \neq 0$, and let $P(z) = \sum_{i=0}^{k-1} a_i z^i$ be the characteristic polynomial associated to recursion (30). Let λ_j, $j = 1 : p$, where $p \leq k$, be the roots of $P(z)$, and let $m(\lambda_j)$ be the multiplicity of λ_j. The general form of the sequence $(x_n)_{n\geq 0}$ satisfying the linear recursion (30) is

$$x_n = \sum_{j=1}^{p} \left(\sum_{i=0}^{m(\lambda_j)-1} C_{i,j} n^i \right) \lambda_j^n, \quad \forall\, n \geq 0, \tag{31}$$

where $C_{i,j}$ are constant numbers.

Example: Find the general formula for the terms of the Fibonacci sequence

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ \ldots,$$

where each new term is the sum of the previous two terms in the sequence.

[2]The same characteristic polynomial corresponds to the linear ODE with constant coefficients $\sum_{i=0}^{k} a_i y^{(i)}(x) = 0$.

Answer: By definition, the terms of the Fibonacci sequence satisfy the linear recursion $x_{n+2} = x_{n+1} + x_n$, for all $n \geq 0$, with $x_0 = 1$, and $x_1 = 1$. This recursion can be written in the form (25) as

$$x_{n+2} - x_{n+1} - x_n \;=\; 0, \;\; \forall\, n \geq 0. \tag{32}$$

The characteristic polynomial associated to the linear recursion (32) is

$$P(z) \;=\; z^2 - z - 1,$$

and the roots of $P(z)$ are

$$\lambda_1 \;=\; \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \lambda_2 \;=\; \frac{1-\sqrt{5}}{2}. \tag{33}$$

From Theorem 0.1, we find that

$$x_n \;=\; C_1\lambda_1^n \;+\; C_2\lambda_2^n, \;\; \forall\, n \geq 0, \tag{34}$$

where λ_1 and λ_2 are given by (33). The constants C_1 and C_2 are chosen in such a way that $x_0 = 1$ and $x_1 = 1$, i.e., such that

$$\begin{cases} C_1 + C_2 \;=\; 1; \\ C_1\lambda_1 + C_2\lambda_2 \;=\; 1. \end{cases}$$

The solution to this linear system is

$$C_1 \;=\; \frac{\sqrt{5}+1}{2\sqrt{5}} \quad \text{and} \quad C_2 \;=\; \frac{\sqrt{5}-1}{2\sqrt{5}}.$$

We conclude from (34) that the general formula for $(x_n)_{n\geq 0}$ is

$$\begin{aligned} x_n \;&=\; \frac{\sqrt{5}+1}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}-1}{2\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n \\ &=\; \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}, \;\; \forall\, n \geq 0. \quad \Box \end{aligned}$$

A complete proof of Theorem 0.1 is technical and beyond the scope of this book. For better understanding, we provide more details for the case when the polynomial $P(z)$ has k different roots, denoted by $\lambda_1, \lambda_2, \ldots \lambda_k$. We want to show that, if the sequence $(x_n)_{n\geq 0}$ satisfies the recursion

$$\sum_{i=0}^{k} a_i x_{n+i} \;=\; 0, \;\; \forall\, n \geq 0, \tag{35}$$

then there exist constants C_j, $j = 1:k$, such that

$$x_n \;=\; \sum_{j=1}^{k} C_j\lambda_j^n, \;\; \forall\, n \geq 0, \tag{36}$$

which is what the general formula (31) reduces to in this case.

If λ_j is a root of $P(z)$, then $P(\lambda_j) = \sum_{i=0}^{k} a_i\, \lambda_j^i = 0$. It is easy to see that the sequence $y_n \;=\; C\lambda_j^n$, $n \geq 0$, where C is an arbitrary constant, satisfies the linear recursion (25):

$$\sum_{i=0}^{k} a_i y_{n+i} \;=\; \sum_{i=0}^{k} a_i\, C\lambda_j^{n+i} \;=\; C\lambda_j^n \sum_{i=0}^{k} a_i\, \lambda_j^i \;=\; C\lambda_j^n\, P(\lambda_j) \;=\; 0.$$

Using the properties (26) and (27), it follows that the sequence $(z_n)_{n\geq 0}$ given by

$$z_n \;=\; \sum_{j=1}^{k} C_j\lambda_j^n, \;\; \forall\, n \geq 0, \tag{37}$$

satisfies the linear recursion (35), where C_j, $j = 1:k$, are arbitrary constants.

Let $(x_n)_{n\geq 0}$ satisfying recursion (35), and let $x_0, x_1, \ldots, x_{k-1}$ be the first k numbers of the sequence. If we can find constants C_j, $j = 1:k$, such that the first k numbers in the sequence $(x_n)_{n\geq 0}$ and in the sequence $(z_n)_{n\geq 0}$ given by (37) are equal, then it is easy to see, e.g., by complete induction, that $x_n = z_n$, for all $n \geq 0$, i.e., that the two sequences are identical, which is what we want to show.

We are looking for constants $(C_j)_{j=1:k}$, such that

$$x_i \;=\; z_i \;=\; \sum_{j=1}^{k} C_j\lambda_j^i, \;\; \forall\, i = 0:(k-1). \tag{38}$$

In other words, C_j, $j = 1:k$, must solve the linear system

$$\sum_{j=1}^{k} C_j\lambda_j^i \;=\; x_i, \;\; \forall\, i = 0:(k-1),$$

which can be written in matrix form as

$$AC \;=\; b, \tag{39}$$

where A is the $k \times k$ matrix given by

$$A \;=\; \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_k \\ \lambda_1^2 & \lambda_2^2 & \ldots & \lambda_k^2 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \ldots & \lambda_k^{k-1} \end{pmatrix}$$

and C and b denote the $k \times 1$ row vectors

$$C = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k-1} \end{pmatrix}.$$

The matrix A is called a Vandermonde matrix. It can be shown (but is not straightforward to prove) that

$$\det(A) = \prod_{1 \le i < j \le k} (\lambda_j - \lambda_i).$$

Since we assumed that the roots $\lambda_1, \lambda_2, \ldots \lambda_k$ of the characteristic polynomial $P(z)$ are all different, we conclude that $\det(A) \neq 0$. Therefore the matrix A is nonsingular and the linear system (39) has a unique solution.

Then, if $(C_j)_{j=1:k}$ represents the unique solution of the linear system (39), the sequence $(x_n)_{n\ge 0}$ given by $x_n = \sum_{j=1}^{k} C_j \lambda_j^n$, for $n \ge 0$, is the only sequence satisfying the linear recursion (35) and having the first k terms equal to x_0, $x_1, \ldots, x_{k-1}$.

0.4 The "Big O" and "little o" notations

The need for the "Big O" notation becomes clear when looking at the behavior of a polynomial $P(x)$ when the argument x is large. Let

$$P(x) = \sum_{k=0}^{n} a_k x^k,$$

with $a_n \neq 0$. It is easy to see that, as $x \to \infty$, the term of largest degree, i.e., $a_n x^n$, dominates all the other terms:

$$\lim_{x\to\infty} \frac{|P(x)|}{x^n} = \lim_{x\to\infty} \frac{\left|\sum_{k=0}^{n} a_k x^k\right|}{x^n} = \lim_{x\to\infty} \left| a_n + \sum_{k=0}^{n-1} \frac{a_k}{x^{n-k}} \right| = |a_n|. \quad (40)$$

The "Big O" notation is used to write the information contained in (40) in a simplified way that is well suited to computations, i.e.,

$$P(x) = O(x^n), \quad \text{as} \quad x \to \infty.$$

Formally, the following definition can be given:

Definition 0.5. *Let $f, g : \mathbb{R} \to \mathbb{R}$. We write that $f(x) = O(g(x))$, as $x \to \infty$, if and only if ("iff") there exist constants $C > 0$ and $M > 0$ such that $\left|\frac{f(x)}{g(x)}\right| \le C$, for any $x \ge M$. This can be written equivalently as*

$$f(x) = O(g(x)), \quad \text{as} \quad x \to \infty, \quad \text{iff} \quad \limsup_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| < \infty. \quad (41)$$

The "Big O" notation can also be used for $x \to -\infty$, as well as for $x \to a$:

$$f(x) = O(g(x)), \quad \text{as} \quad x \to -\infty, \quad \text{iff} \quad \limsup_{x\to-\infty} \left|\frac{f(x)}{g(x)}\right| < \infty; \quad (42)$$

$$f(x) = O(g(x)), \quad \text{as} \quad x \to a, \quad \text{iff} \quad \limsup_{x\to a} \left|\frac{f(x)}{g(x)}\right| < \infty. \quad (43)$$

We note that the estimate (43) is most often used for $a = 0$, i.e., for $x \to 0$.

The "little o" notation refers to functions whose ratios tend to 0 at certain points, and can be defined for $x \to \infty$, $x \to a$, and $x \to -\infty$ as follows:

Definition 0.6. *Let $f, g : \mathbb{R} \to \mathbb{R}$. Then,*

$$f(x) = o(g(x)), \quad \text{as} \quad x \to \infty, \quad \text{iff} \quad \lim_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| = 0; \quad (44)$$

$$f(x) = o(g(x)), \quad \text{as} \quad x \to -\infty, \quad \text{iff} \quad \lim_{x\to-\infty} \left|\frac{f(x)}{g(x)}\right| = 0; \quad (45)$$

$$f(x) = o(g(x)), \quad \text{as} \quad x \to a, \quad \text{iff} \quad \lim_{x\to a} \left|\frac{f(x)}{g(x)}\right| = 0. \quad (46)$$

The rules for operations with $O(\cdot)$ and $o(\cdot)$, e.g., for additions, subtractions, multiplications follow from the definitions (41)–(46).

Example: If $0 < n < m$, then

$$x^n = o(x^m), \quad \text{as} \quad x \to \infty; \quad (47)$$

$$O(x^n) + O(x^m) = O(x^m), \quad \text{as} \quad x \to \infty; \quad (48)$$

$$x^m = o(x^n), \quad \text{as} \quad x \to 0; \quad (49)$$

$$O(x^n) + O(x^m) = O(x^n), \quad \text{as} \quad x \to 0. \quad (50)$$

Answer: We only sketch the proofs of (47–49).

To prove (47), note that, since $m > n$,

$$\lim_{x\to\infty} \left|\frac{x^n}{x^m}\right| = \lim_{x\to\infty} \frac{1}{x^{m-n}} = 0.$$

Therefore, $x^n = o(x^m)$, as $x \to \infty$; cf. definition (44).

To prove (49), we obtain similarly that

$$\lim_{x\to 0}\left|\frac{x^m}{x^n}\right| = \lim_{x\to 0}|x|^{m-n} = 0,$$

and therefore $x^m = o(x^n)$, as $x \to 0$; cf. definition (46).

To prove (48), i.e., that $O(x^n)+O(x^m) = O(x^m)$, as $x \to \infty$, if $0 < n < m$, let $f, g : \mathbb{R} \to \mathbb{R}$ such that $f(x) = O(x^n)$, as $x \to \infty$, and $g(x) = O(x^m)$, as $x \to \infty$. By definition (41), it follows that

$$\limsup_{x\to\infty}\left|\frac{f(x)}{x^n}\right| < \infty \quad \text{and} \quad \limsup_{x\to\infty}\left|\frac{g(x)}{x^m}\right| < \infty.$$

In other words, there exist constants C_f, M_f and C_g, M_g, such that

$$\left|\frac{f(x)}{x^n}\right| \le C_f, \ \forall\, x \ge M_f \quad \text{and} \quad \left|\frac{g(x)}{x^m}\right| \le C_g, \ \forall\, x \ge M_g. \qquad (51)$$

Let $h(x) = f(x) + g(x)$. To show that $h(x) = O(x^m)$, as $x \to \infty$, it is enough to prove that there exist constants C_h and M_h such that

$$\left|\frac{h(x)}{x^m}\right| \le C_h, \ \forall\, x \ge M_h. \qquad (52)$$

From (51), it follows that, for any $x \ge \max(M_f, M_g)$,

$$\left|\frac{h(x)}{x^m}\right| = \left|\frac{f(x)+g(x)}{x^m}\right| \le \left|\frac{f(x)}{x^m}\right| + \left|\frac{g(x)}{x^m}\right| \le \frac{1}{x^{m-n}}\,C_f + C_g. \quad (53)$$

Note that $\lim_{x\to\infty}\frac{1}{x^{m-n}} = 0$, since $m > n$. From (53), it follows that we can find constants C_h and M_h such that (52) holds true, and therefore (48) is proved. $\square$

Similarly, it can be shown that, for any $n > 0$,

$$\begin{aligned} O(x^n) + O(x^n) &= O(x^n); & O(x^n) - O(x^n) &= O(x^n); \\ o(x^n) + o(x^n) &= o(x^n); & o(x^n) - o(x^n) &= o(x^n). \end{aligned}$$

Finally, note that, by definition, $-O(g(x)) = O(g(x))$, and, similarly, $-o(g(x)) = o(g(x))$. More generally, for any constant $c \neq 0$, we can write that

$$\begin{aligned} O(cg(x)) &= O(g(x)) \ \text{and} \ c\,O(g(x)) = O(g(x)); \\ o(cg(x)) &= o(g(x)) \ \text{and} \ c\,o(g(x)) = o(g(x)). \end{aligned}$$

The $O(\cdot)$ and $o(\cdot)$ notations are useful for Taylor approximations as well as for finite difference approximations; see, sections 5.1 and 6.1 for details.

0.5 Exercises

1. Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function.

   (i) Show that $xf(x)$ is an even function and $x^2 f(x)$ is an odd function.

   (ii) Show that the function $g_1 : \mathbb{R} \to \mathbb{R}$ given by $g_1(x) = f(x^2)$ is an even function and that the function $g_2 : \mathbb{R} \to \mathbb{R}$ given by $g_2(x) = f(x^3)$ is an odd function.

   (iii) Let $h : \mathbb{R} \to \mathbb{R}$ be defined as $h(x) = x^i f(x^j)$, where i and j are positive integers. When is $h(x)$ an odd function?

2. Let $S(n,2) = \sum_{k=1}^n k^2$ and $S(n,3) = \sum_{k=1}^n k^3$.

   (i) Let $T(n,2,x) = \sum_{k=1}^n k^2\, x^k$. Use formula (23) for $j = 1$, i.e.,

   $$T(n,2,x) = x\,\frac{d}{dx}(T(n,1,x)),$$

   and formula (24) for $T(n,1,x)$, to show that

   $$T(n,2,x) = \frac{x + x^2 - (n+1)^2 x^{n+1} + (2n^2+2n-1)x^{n+2} - n^2 x^{n+3}}{(1-x)^3}.$$

   (ii) Note that $S(n,2) = T(n,2,1)$. Use l'Hôpital's rule to evaluate $T(n,2,1)$, and conclude that $S(n,2) = \frac{n(n+1)(2n+1)}{6}$.

   (iii) Compute $T(n,3,x) = \sum_{k=1}^n k^3\, x^k$ using formula (23) for $j = 2$, i.e,

   $$T(n,3,x) = x\,\frac{d}{dx}(T(n,2,x)).$$

   (iv) Note that $S(n,3) = T(n,3,1)$. Use l'Hôpital's rule to evaluate $T(n,3,1)$, and conclude that $S(n,3) = \left(\frac{n(n+1)}{2}\right)^2$.

3. Compute $S(n,4) = \sum_{k=1}^n k^4$ using the recursion formula (19) for $i = 4$, the fact that $S(n,0) = n$, and formulas (11–13) for $S(n,1)$, $S(n,2)$, and $S(n,3)$.

4. It is easy to see that the sequence $(x_n)_{n\ge 1}$ given by $x_n = \sum_{k=1}^n k^2$ satisfies the recursion

   $$x_{n+1} = x_n + (n+1)^2, \ \forall\, n \ge 1, \qquad (54)$$

with $x_1 = 1$.

(i) By substituting $n+1$ for n in (54), obtain that

$$x_{n+2} = x_{n+1} + (n+2)^2. \tag{55}$$

Subtract (54) from (55) to find that

$$x_{n+2} = 2x_{n+1} - x_n + 2n + 3, \quad \forall\, n \geq 1, \tag{56}$$

with $x_1 = 1$ and $x_2 = 5$.

(ii) Similarly, substitute $n+1$ for n in (56) and obtain that

$$x_{n+3} = 2x_{n+2} - x_{n+1} + 2(n+1) + 3. \tag{57}$$

Subtract (56) from (57) to find that

$$x_{n+3} = 3x_{n+2} - 3x_{n+1} + x_n + 2, \quad \forall\, n \geq 1,$$

with $x_1 = 1$, $x_2 = 5$, and $x_3 = 14$.

(iii) Use a similar method to prove that the sequence $(x_n)_{n\geq 0}$ satisfies the linear recursion

$$x_{n+4} - 4x_{n+3} + 6x_{n+2} - 4x_{n+1} + x_n = 0, \quad \forall\, n \geq 1. \tag{58}$$

The characteristic polynomial associated to the recursion (58) is

$$P(z) = z^4 - 4z^3 + 6z^2 - 4z + 1 = (z-1)^4.$$

Use the fact that $x_1 = 1$, $x_2 = 5$, $x_3 = 14$, and $x_4 = 30$ to show that

$$x_n = \frac{n(n+1)(2n+1)}{6}, \quad \forall\, n \geq 1,$$

and conclude that

$$S(n,2) = \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \forall\, n \geq 1.$$

5. Find the general form of the sequence $(x_n)_{n\geq 0}$ satisfying the linear recursion

$$x_{n+3} = 2x_{n+1} + x_n, \quad \forall\, n \geq 0,$$

with $x_0 = 1$, $x_1 = -1$, and $x_2 = 1$.

6. The sequence $(x_n)_{n\geq 0}$ satisfies the recursion

$$x_{n+1} = 3x_n + 2, \quad \forall\, n \geq 0,$$

with $x_0 = 1$.

(i) Show that the sequence $(x_n)_{n\geq 0}$ satisfies the linear recursion

$$x_{n+2} = 4x_{n+1} - 3x_n, \quad \forall\, n \geq 0,$$

with $x_0 = 1$ and $x_1 = 5$.

(ii) Find the general formula for x_n, $n \geq 0$.

7. The sequence $(x_n)_{n\geq 0}$ satisfies the recursion

$$x_{n+1} = 3x_n + n + 2, \quad \forall\, n \geq 0,$$

with $x_0 = 1$.

(i) Show that the sequence $(x_n)_{n\geq 0}$ satisfies the linear recursion

$$x_{n+3} = 5x_{n+2} - 7x_{n+1} + 3x_n, \quad \forall\, n \geq 0,$$

with $x_0 = 1$, $x_1 = 5$, and $x_2 = 18$.

(ii) Find the general formula for x_n, $n \geq 0$.

8. Let $P(z) = \sum_{i=0}^{k} a_i z^i$ be the characteristic polynomial corresponding to the linear recursion

$$\sum_{i=0}^{k} a_i x_{n+i} = 0, \quad \forall\, n \geq 0. \tag{59}$$

Assume that λ is a root of multiplicity 2 of $P(z)$. Show that the sequence $(y_n)_{n\geq 0}$ given by

$$y_n = Cn\lambda^n, \quad n \geq 0,$$

where C is an arbitrary constant, satisfies the recursion (59).

Hint: Show that

$$\sum_{i=0}^{k} a_i y_{n+i} = Cn\lambda^n P(\lambda) + C\lambda^{n+1} P'(\lambda), \quad \forall\, n \geq 0,$$

and recall that λ is a root of multiplicity 2 of the polynomial $P(z)$ if and only if $P(\lambda) = 0$ and $P'(\lambda) = 0$.

9. Let $n > 0$. Show that

$$O(x^n) + O(x^n) = O(x^n), \quad \text{as} \quad x \to 0; \tag{60}$$
$$o(x^n) + o(x^n) = o(x^n), \quad \text{as} \quad x \to 0. \tag{61}$$

For example, to prove (60), let $f(x) = O(x^n)$ and $g(x) = O(x^n)$ as $x \to 0$, and show that $f(x) + g(x) = O(x^n)$ as $x \to 0$, i.e., that

$$\limsup_{x\to 0} \left| \frac{f(x)+g(x)}{x^n} \right| < \infty.$$

10. Prove that

$$\sum_{k=1}^{n} k^2 = O(n^3), \quad \text{as} \quad n \to \infty;$$
$$\sum_{k=1}^{n} k^2 = \frac{n^3}{3} + O(n^2), \quad \text{as} \quad n \to \infty,$$

i.e., show that

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^{n} k^2}{n^3} < \infty$$

and that

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^{n} k^2 - \frac{n^3}{3}}{n^2} < \infty.$$

Similarly, prove that

$$\sum_{k=1}^{n} k^3 = O(n^4), \quad \text{as} \quad n \to \infty;$$
$$\sum_{k=1}^{n} k^3 = \frac{n^4}{4} + O(n^3), \quad \text{as} \quad n \to \infty.$$

Chapter 1

Calculus review. Plain vanilla options.

Brief review of differentiation: Product Rule, Quotient Rule, Chain Rule for functions of one variable. Derivative of the inverse function.

Brief review of integration: Fundamental Theorem of Calculus, integration by parts, integration by substitution.

Differentiating definite integrals with respect to parameters in the limits of integration and with respect to parameters in the integrated function.

Limits. L'Hôpital's Rule. Connections to Taylor expansions.

Multivariable functions. Partial derivatives. Gradient and Hessian of multivariable functions.

1.1 Brief review of differentiation

We begin by briefly reviewing elementary differentiation topics for functions of one variable.

The function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at the point $x \in \mathbb{R}$ if the limit

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists, in which case the derivative[1] $f'(x)$ is defined as

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}. \tag{1.1}$$

The function $f(x)$ is called differentiable if it is differentiable at all points x.

Theorem 1.1. ***(Product Rule.)*** *The product $f(x)g(x)$ of two differentiable functions $f(x)$ and $g(x)$ is differentiable, and*

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x). \tag{1.2}$$

[1] We anticipate by noting that the forward and backward finite difference approximations of the first derivative of a function can be obtained from definition (1.1); see (6.3) and (6.5).

Theorem 1.2. ***(Quotient Rule.)*** *The quotient $\frac{f(x)}{g(x)}$ of two differentiable functions $f(x)$ and $g(x)$ is differentiable at every point x where the function $\frac{f(x)}{g(x)}$ is well defined, and*

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2} \tag{1.3}$$

Theorem 1.3. ***(Chain Rule.)*** *The composite function $(g \circ f)(x) = g(f(x))$ of two differentiable functions $f(x)$ and $g(x)$ is differentiable at every point x where $g(f(x))$ is well defined, and*

$$(g(f(x)))' = g'(f(x))\, f'(x). \tag{1.4}$$

The Chain Rule for multivariable functions is presented in section 7.1.

The Chain Rule formula (1.4) can also be written as

$$\frac{dg}{dx} = \frac{dg}{du}\frac{du}{dx},$$

where $u = f(x)$ is a function of x and $g = g(u) = g(f(x))$.

Example: Chain Rule is often used for power functions, exponential functions, and logarithmic functions:

$$\frac{d}{dx}((f(x))^n) = n(f(x))^{n-1}\, f'(x); \tag{1.5}$$

$$\frac{d}{dx}\left(e^{f(x)}\right) = e^{f(x)}\, f'(x); \tag{1.6}$$

$$\frac{d}{dx}(\ln f(x)) = \frac{f'(x)}{f(x)}. \quad \square \tag{1.7}$$

The derivative of the inverse of a function is computed as follows:

Lemma 1.1. *Let $f : [a,\ b] \to [c,\ d]$ be a differentiable function, and assume that $f(x)$ has an inverse function denoted by $f^{-1}(x)$, with $f^{-1} : [c,\ d] \to [a,\ b]$. The function $f^{-1}(x)$ is differentiable at every point $x \in [c,\ d]$ where $f'(f^{-1}(x)) \neq 0$ and*

$$\left(f^{-1}(x)\right)' = \frac{1}{f'(f^{-1}(x))}. \tag{1.8}$$

While we do not prove here that $f^{-1}(x)$ is differentiable (this can be done, e.g., by using the definition (1.1) of the derivative of a function), we derive formula (1.8). Recall from (1.4) that

$$(g(f(z)))' = g'(f(z))\, f'(z). \tag{1.9}$$

Let $g = f^{-1}$ in (1.9). Since $g(f(z)) = f^{-1}(f(z)) = z$, it follows that

$$1 = \left(f^{-1}\right)'(f(z)) \cdot f'(z). \tag{1.10}$$

Let $z = f^{-1}(x)$ in (1.10). Then, $f(z) = f(f^{-1}(x)) = x$ and (1.10) becomes

$$1 = \left(f^{-1}\right)'(x) \cdot f'(f^{-1}(x)) = \left(f^{-1}(x)\right)' \cdot f'(f^{-1}(x)). \tag{1.11}$$

If $f'(f^{-1}(x)) \neq 0$, formula (1.8) follows immediately from (1.11).

Examples:

$$\begin{aligned}
\frac{d}{dx}\left(x\, e^{3x^2-1}\right) &= (1+6x^2)\, e^{3x^2-1}; \\
\frac{d}{dx}\left(\frac{\sqrt{3x^2-1}}{\sqrt{3x^2-1}+4}\right) &= \frac{d}{dx}\left(1 - \frac{4}{\sqrt{3x^2-1}+4}\right); \\
&= \frac{12x}{\sqrt{3x^2-1}\,(\sqrt{3x^2-1}+4)^2} \\
\frac{d}{dx}\left(\frac{e^{x^2}+1}{x-1}\right) &= \frac{(2x^2-2x-1)e^{x^2}-1}{(x-1)^2}; \\
\frac{d}{dx}\left(\ln\left(\frac{x}{2x^2+1}\right)\right) &= \frac{d}{dx}\left(\ln(x) - \ln(2x^2+1)\right) = \frac{-2x^2+1}{x(2x^2+1)}. \quad \square
\end{aligned}$$

1.2 Brief review of integration

In this section, we briefly review several elementary integration topics, both for antiderivatives and for definite integrals.

Let $f : \mathbb{R} \to \mathbb{R}$ be an integrable function[2]. Recall that $F(x)$ is the antiderivative of $f(x)$ if and only if $F'(x) = f(x)$, i.e.,

$$F(x) = \int f(x)\, dx \iff F'(x) = f(x). \tag{1.12}$$

The Fundamental Theorem of Calculus provides a formula for evaluating the definite integral of a continuous function, if a closed formula for its antiderivative is known.

[2]Throughout the book, by integrable function we mean Riemann integrable.

Theorem 1.4. ***(Fundamental Theorem of Calculus.)*** *Let $f(x)$ be a continuous function on the interval $[a,b]$, and let $F(x)$ be the antiderivative of $f(x)$. Then*

$$\int_a^b f(x)\,dx \;=\; F(x)|_a^b \;=\; F(b)-F(a).$$

Integration by parts is the counterpart for integration of the product rule.

Theorem 1.5. ***(Integration by parts.)*** *Let $f(x)$ and $g(x)$ be continuous functions. Then*

$$\int f(x)g(x)\,dx \;=\; F(x)g(x) \;-\; \int F(x)g'(x)\,dx, \tag{1.13}$$

where $F(x)=\int f(x)dx$ is the antiderivative of $f(x)$. For definite integrals,

$$\int_a^b f(x)g(x)\,dx \;=\; F(b)g(b)-F(a)g(a) \;-\; \int_a^b F(x)g'(x)\,dx. \tag{1.14}$$

Proof. We apply the product rule (1.2) to the function $F(x)g(x)$ and obtain

$$(F(x)g(x))' \;=\; (F(x))'g(x)+F(x)g'(x) \;=\; f(x)g(x)+F(x)g'(x), \tag{1.15}$$

since $F'(x)=f(x)$; cf. (1.12). By taking antiderivatives in (1.15), we conclude that

$$F(x)g(x) \;=\; \int f(x)g(x)\,dx \;+\; \int F(x)g'(x)\,dx,$$

which is equivalent to (1.13).

To derive the formula (1.14) for definite integrals, we apply the Fundamental Theorem of Calculus to (1.15) and obtain that

$$\int_a^b (f(x)g(x)+F(x)g'(x))\,dx \;=\; \int_a^b (F(x)g(x))'\,dx \;=\; (F(x)g(x))\,|_a^b\,.$$

This can be written as

$$\int_a^b f(x)g(x)\,dx \;+\; \int_a^b F(x)g'(x)\,dx \;=\; F(b)g(b)-F(a)g(a),$$

which is equivalent to (1.14). □

Integration by substitution is the counterpart for integration of the chain rule.

Theorem 1.6. ***(Integration by substitution.)*** *Let $f(x)$ be an integrable function. Assume that $g(u)$ is an invertible and continuously differentiable function. The substitution $x=g(u)$ changes the integration variable from x to u as follows:*

$$\int f(x)\,dx \;=\; \int f(g(u))g'(u)\,du. \tag{1.16}$$

For definite integrals,

$$\int_a^b f(x)\,dx \;=\; \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)\,du. \tag{1.17}$$

Informally, formula (1.16) follows from the fact that, by differentiating the substitution formula $x=g(u)$, it follows that $dx=g'(u)du$. The bounds for the definite integrals in (1.17) change according to the rule $u=g^{-1}(x)$. In other words, $x=a$ and $x=b$ correspond to $u=g^{-1}(a)$ and $u=g^{-1}(b)$, respectively. Formal proofs of these results are given below.

Proof. Let $F(x)=\int f(x)dx$ be the antiderivative of $f(x)$. The chain rule (1.4) applied to $F(g(u))$ yields

$$(F(g(u)))' \;=\; F'(g(u))\;g'(u) \;=\; f(g(u))\;g'(u), \tag{1.18}$$

since $F'=f$; cf. (1.12). Integrating (1.18) with respect to u, we find that

$$\int f(g(u))g'(u)\,du \;=\; \int (F(g(u)))'\,du \;=\; F(g(u)). \tag{1.19}$$

Using the substitution $x=g(u)$ we notice that

$$F(g(u)) \;=\; F(x) \;=\; \int f(x)\,dx. \tag{1.20}$$

From (1.19) and (1.20), we conclude that

$$\int f(x)\,dx \;=\; \int f(g(u))g'(u)\,du.$$

To obtain the formula (1.17) for definite integrals, we apply the Fundamental Theorem of Calculus to (1.18) and obtain

$$\begin{aligned}\int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)\,du \;&=\; \int_{g^{-1}(a)}^{g^{-1}(b)} (F(g(u)))'\,du \;=\; F(g(u))|_{g^{-1}(a)}^{g^{-1}(b)}\\ &=\; F(g(g^{-1}(b)))-F(g(g^{-1}(a)))\\ &=\; F(b)-F(a).\end{aligned} \tag{1.21}$$

From the Fundamental Theorem of Calculus, we find that

$$\int_a^b f(x)\,dx = F(b) - F(a), \tag{1.22}$$

since $F(x) = \int f(x)dx$. From (1.21) and (1.22), we conclude that

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)\,du.$$

□

We note that, while product rule and chain rule correspond to integration by parts and integration by substitution, the quotient rule does not have a counterpart in integration.

Examples:

$$\int \ln(1+x)\,dx = (1+x)\,\ln(1+x) - x + C$$

(Integration by parts: $f(x) = 1;\ F(x) = 1 + x;\ g(x) = \ln(1+x)$);

$$\int_1^3 xe^x\,dx = (xe^x - e^x)\big|_1^3 = 2e^3$$

(Integration by parts: $f(x) = e^x;\ F(x) = e^x;\ g(x) = x$);

$$\int x^2 \ln(x)\,dx = \frac{x^3\ln(x)}{3} - \frac{x^3}{9} + C$$

(Integration by parts: $f(x) = x^2;\ F(x) = \frac{x^3}{3};\ g(x) = \ln(x)$);

$$\int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx = 2e^{\sqrt{x}} + C;\quad \text{Substitution: } u = \sqrt{x};$$

$$\int_{-1}^0 x^2(x^3-1)^4\,dx = \left.\frac{(x^3-1)^5}{15}\right|_{-1}^0 = \frac{31}{15};\quad \text{Substitution: } u = x^3 - 1;$$

$$\int \frac{e^x + e^{-x}}{e^x - e^{-x}}\,dx = \ln(e^x - e^{-x}) + C;\quad \text{Substitution:} u = e^x - e^{-x}.\Box$$

1.3 Differentiating definite integrals

A definite integral of the form $\int_a^b f(x)\,dx$ is a real number. However, if a definite integral has functions as limits of integration, e.g.,

$$\int_{a(t)}^{b(t)} f(x)\,dx,$$

or if the function to be integrated is a function of the integrating variable and of another variable, e.g.,

$$\int_a^b f(x,t)\,dx$$

then the result of the integration is a function (of the variable t in both cases above). If certain conditions are met, this function is differentiable.

Lemma 1.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then,*

$$\frac{d}{dt}\left(\int_{a(t)}^{b(t)} f(x)\,dx\right) = f(b(t))b'(t) - f(a(t))a'(t), \tag{1.23}$$

where $a(t)$ and $b(t)$ are differentiable functions.

Proof. Let $F(x) = \int f(x)dx$ be the antiderivative of $f(x)$. Define the function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(t) = \int_{a(t)}^{b(t)} f(x)\,dx.$$

From the Fundamental Theorem of Calculus, see Theorem 1.4, it follows that

$$g(t) = F(b(t)) - F(a(t)).$$

Recall that $F'(x) = f(x)$. Then $g(t)$ is a differentiable function, since $a(t)$ and $b(t)$ are differentiable. Using chain rule (1.4), we find that

$$g'(t) = F'(b(t))b'(t) - F'(a(t))a'(t) = f(b(t))b'(t) - f(a(t))a'(t).$$

□

Lemma 1.3. *Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that the partial derivative $\frac{\partial f}{\partial t}(x,t)$ exists*[3] *and is continuous in both variables x and t. Then,*

$$\frac{d}{dt}\left(\int_a^b f(x,t)\,dx\right) = \int_a^b \frac{\partial f}{\partial t}(x,t)\,dx.$$

A rigorous proof of this lemma can be given by introducing the function

$$g(t) = \int_a^b f(x,t)\,dx$$

and using definition (1.1) of the derivative of a function to compute $g'(t)$, i.e.,

$$g'(t) = \lim_{h\to 0}\frac{g(t+h) - g(t)}{h} = \lim_{h\to 0}\int_a^b \frac{f(x,t+h) - f(x,t)}{h}\,dx.$$

For our purposes, it is enough to use Lemma 1.3 without studying its proof.

[3]For details on partial derivatives of functions of two variables, see section 1.6.1.

Lemma 1.4. *Let $f(x,t)$ be a continuous function such that the partial derivative $\frac{\partial f}{\partial t}(x,t)$ exists and is continuous. Then,*

$$\frac{d}{dt}\left(\int_{a(t)}^{b(t)} f(x,t)\,dx\right) = \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}(x,t)\,dx + f(b(t),t)b'(t) - f(a(t),t)a'(t).$$

Note that Lemma 1.2 and Lemma 1.3 are special cases of Lemma 1.4.

1.4 Limits

Definition 1.1. *Let $g : \mathbb{R} \to \mathbb{R}$. The limit of $g(x)$ as $x \to x_0$ exists and is finite and equal to l if and only if for any $\epsilon > 0$ there exists $\delta > 0$ such that $|g(x) - l| < \epsilon$ for all $x \in (x_0 - \delta,\ x_0 + \delta)$, i.e.,*

$$\lim_{x\to x_0} g(x) = l \quad \text{iff} \quad \forall\, \epsilon > 0\ \exists\, \delta > 0 \ \text{ such that } \ |g(x) - l| < \epsilon,\ \forall\, |x - x_0| < \delta.$$

Similarly,

$$\begin{aligned}
\lim_{x\to x_0} g(x) = \infty \quad &\text{iff} \ \ \forall\, C > 0\ \exists\, \delta > 0 \ \text{ such that } \ g(x) > C,\ \forall\, |x - x_0| < \delta;\\
\lim_{x\to x_0} g(x) = -\infty \quad &\text{iff} \ \ \forall\, C < 0\ \exists\, \delta > 0 \ \text{ such that } \ g(x) < C,\ \forall\, |x - x_0| < \delta.
\end{aligned}$$

Limits are used, for example, to define the derivative of a function; cf. (1.1).

In this book, we will rarely need to use Definition 1.1 to compute the limit of a function. We note that many limits can be computed by using the fact that, at infinity, exponential functions are much bigger that absolute values of polynomials, which are in turn much bigger than logarithms.

Theorem 1.7. *If $P(x)$ and $Q(x)$ are polynomials and $c > 1$ is a fixed constant, then*

$$\lim_{x\to\infty} \frac{P(x)}{c^x} = 0, \ \ \forall\ c > 1; \tag{1.24}$$

$$\lim_{x\to\infty} \frac{\ln|Q(x)|}{P(x)} = 0. \tag{1.25}$$

Examples: From (1.24) and (1.25), it is easy to see that

$$\lim_{x\to\infty} x^5\, e^{-x} = 0; \quad \lim_{x\to\infty} \frac{\ln(x)}{x^3} = 0. \quad \square$$

A general method to prove (1.24) and (1.25), as well as computing many other limits, is to show that the function whose limit is to be computed is

either increasing or decreasing, which means that the limit exists, and use Definition 1.1 to compute it. Such formal proofs are beyond our scope here.

In the course of the material presented in this book, we will use several limits that are simple consequences of (1.24) and (1.25).

Lemma 1.5. *Let $c > 0$ be a positive constant. Then,*

$$\lim_{x\to\infty} x^{\frac{1}{x}} = 1; \tag{1.26}$$

$$\lim_{x\to\infty} c^{\frac{1}{x}} = 1; \tag{1.27}$$

$$\lim_{x\searrow 0} x^x = 1, \tag{1.28}$$

where the notation $x \searrow 0$ means that x goes to 0 while always being positive, i.e., $x \to 0$ with $x > 0$.

Proof. We only prove (1.26); the other limits can be obtained similarly. We compute the limit of the logarithm of $x^{\frac{1}{x}}$ as $x \to \infty$. Using (1.25), we find that

$$\lim_{x\to\infty} \ln\left(x^{\frac{1}{x}}\right) = \lim_{x\to\infty} \frac{\ln(x)}{x} = 0,$$

and therefore

$$\lim_{x\to\infty} x^{\frac{1}{x}} = \lim_{x\to\infty} \exp\left(\ln\left(x^{\frac{1}{x}}\right)\right) = \exp(0) = 1,$$

where $\exp(z) = e^z$. $\square$

We will also use limits of the form (1.26) and (1.27) in the discrete setting of integers k going to ∞.

Lemma 1.6. *If k is a positive integer number, and if $c > 0$ is a positive fixed constant, then*

$$\lim_{k\to\infty} k^{\frac{1}{k}} = 1; \tag{1.29}$$

$$\lim_{k\to\infty} c^{\frac{1}{k}} = 1; \tag{1.30}$$

$$\lim_{k\to\infty} \frac{c^k}{k!} = 0, \tag{1.31}$$

where $k! = 1 \cdot 2 \cdot \ldots \cdot k$.

We conclude by recalling that

$$\lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^x = e, \tag{1.32}$$

which is one possible way to define the number $e \approx 2.71828$.

1.5 L'Hôpital's rule and connections to Taylor expansions

L'Hôpital's rule is a method to compute limits when direct computation would give an undefined result of form $\frac{0}{0}$. Informally, if $\lim_{x\to x_0} f(x) = 0$ and $\lim_{x\to x_0} g(x) = 0$, then $\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}$. Formally, l'Hôpital's rule can be stated as follows:

Theorem 1.8. ***(L'Hôpital's Rule.)*** *Let x_0 be a real number; allow $x_0 = \infty$ and $x_0 = -\infty$ as well. Let $f(x)$ and $g(x)$ be two differentiable functions.*

(i) Assume that $\lim_{x\to x_0} f(x) = 0$ and $\lim_{x\to x_0} g(x) = 0$. If $\lim_{x\to x_0} \frac{f'(x)}{g'(x)}$ exists, and if there exists an interval $(a,\ b)$ around x_0 such that $g'(x) \neq 0$ for all $x \in (a,\ b) \setminus 0$, then the limit $\lim_{x\to x_0} \frac{f(x)}{g(x)}$ also exists and

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} \;=\; \lim_{x\to x_0} \frac{f'(x)}{g'(x)}.$$

(ii) Assume that $\lim_{x\to x_0} f(x)$ is either $-\infty$ or ∞, and that $\lim_{x\to x_0} g(x)$ is either $-\infty$ or ∞. If the limit $\lim_{x\to x_0} \frac{f'(x)}{g'(x)}$ exists, and if there exists an interval $(a,\ b)$ around x_0 such that $g'(x) \neq 0$ for all $x \in (a,\ b) \setminus 0$, then the limit $\lim_{x\to x_0} \frac{f(x)}{g(x)}$ also exists and

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} \;=\; \lim_{x\to x_0} \frac{f'(x)}{g'(x)}.$$

Note that, if $x_0 = -\infty$ of if $x_0 = \infty$ the interval $(a,\ b)$ from Theorem 1.8 is of the form $(-\infty,\ b)$ and $(a,\ \infty)$, respectively.

L'Hôpital's rule can also be applied to other undefined limits such as

$$0\cdot\infty, \quad \frac{1}{0\cdot\infty}, \quad 0^0, \quad \infty^0, \quad \text{and} \quad 1^\infty.$$

In section 5.3, we present linear and quadratic Taylor expansions for several elementary functions; see (5.15–5.24). It is interesting to note that l'Hôpital's rule can be used to prove that these expansions hold true on small intervals. For example, the linear expansion (5.15) of the function e^x around the point 0 is $e^x \approx 1 + x$. Using l'Hôpital's rule, we can show that

$$\lim_{x\to 0} \frac{e^x \;-\; (1+x)}{x^2} \;=\; \frac{1}{2}, \tag{1.33}$$

which means that $e^x \approx 1 + x$ for small values of x, and the approximation is of order 2, i.e.,

$$e^x \;=\; 1 \;+\; x \;+\; O(x^2), \quad \text{as } x \to 0;$$

see (41) for the definition of the $O(\cdot)$ notation.

To prove (1.33), differentiate both the numerator and denominator and obtain the following limit to compute:

$$\lim_{x\to 0} \frac{e^x - 1}{2x}. \tag{1.34}$$

This limit is $\frac{0}{0}$. We attempt to apply l'Hôpital's rule to compute (1.34). By differentiating the numerator and denominator of (1.34), we find that

$$\lim_{x\to 0} \frac{e^x}{2} \;=\; \frac{1}{2}.$$

Then, from l'Hôpital's rule, we obtain that

$$\lim_{x\to 0} \frac{e^x - 1}{2x} \;=\; \frac{1}{2},$$

and, applying again l'Hôpital's rule, we conclude that

$$\lim_{x\to 0} \frac{e^x \;-\; (1+x)}{x^2} \;=\; \frac{1}{2}.$$

1.6 Multivariable functions

Until now, we only considered functions $f(x)$ of one variable. In this section, we introduce functions of several variables, either taking values in the one–dimensional space $\mathbb{R}$, i.e., scalar valued multivariable functions, or taking values in the m–dimensional space $\mathbb{R}^m$, i.e., vector valued multivariable functions.

Scalar Valued Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables denoted by $x_1,\ x_2,\ \ldots,\ x_n$, and let $x = (x_1, x_2, \ldots, x_n)$.

Definition 1.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$. The partial derivative of the function $f(x)$ with respect to the variable x_i is denoted by $\frac{\partial f}{\partial x_i}(x)$ and is defined as*

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h\to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h}, \tag{1.35}$$

if the limit from (1.35) exists and is finite.

In practice, the partial derivative $\frac{\partial f}{\partial x_i}(x)$ is computed by considering the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ to be fixed, and differentiating $f(x)$ as a function of one variable x_i.

A compact formula for (1.35) can be given as follows: Let e_i be the vector with all entries equal to 0 with the exception of the i-th entry, which is equal to 1, i.e., $e_i(j) = 0$, for $j \neq i$, $1 \leq j \leq n$, and $e_i(j) = 1$. Then,

$$\frac{\partial f}{\partial x_i}(x) \;=\; \lim_{h \to 0} \frac{f(x + he_i) \;-\; f(x)}{h}.$$

Partial derivatives of higher order are defined similarly. For example, the second order partial derivative of $f(x)$ first with respect to x_i and then with respect to x_j, with $j \neq i$, is denoted by $\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ and is equal to

$$\frac{\partial^2 f}{\partial x_j\, \partial x_i}(x) \;=\; \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(x)\right),$$

while the second and third partial derivatives of $f(x)$ with respect to x_i are denoted by $\frac{\partial^2 f}{\partial x_i^2}(x)$ and $\frac{\partial^3 f}{\partial x_i^3}(x)$, respectively, and are given by

$$\begin{aligned} \frac{\partial^2 f}{\partial x_i^2}(x) \;&=\; \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}(x)\right); \\ \frac{\partial^3 f}{\partial x_i^3} \;&=\; \frac{\partial}{\partial x_i}\left(\frac{\partial^2 f}{\partial x_i^2}(x)\right). \end{aligned}$$

While the order in which the partial derivatives of a given function are computed might make a difference, i.e., the partial derivative of $f(x)$ first with respect to x_i and then with respect to x_j, with $j \neq i$, is not necessarily equal to the partial derivative of $f(x)$ first with respect to x_j and then with respect to x_i, this is not the case if a function is smooth enough:

Theorem 1.9. *If all the partial derivatives of order k of the function $f(x)$ exist and are continuous, then the order in which partial derivatives of $f(x)$ of order at most k is computed does not matter.*

Definition 1.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables and assume that $f(x)$ is differentiable with respect to all variables x_i, $i = 1 : n$. The gradient $Df(x)$ of the function $f(x)$ is the following row vector of size n:*

$$Df(x) \;=\; \left(\frac{\partial f}{\partial x_1}(x) \;\; \frac{\partial f}{\partial x_2}(x) \;\; \ldots \;\; \frac{\partial f}{\partial x_n}(x)\right). \tag{1.36}$$

Definition 1.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables. The Hessian of $f(x)$ is denoted by $D^2 f(x)$ and is defined as the following $n \times n$ matrix:*

$$D^2 f(x) \;=\; \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \ldots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}. \tag{1.37}$$

Another commonly used notations for the gradient and Hessian of $f(x)$ are $\nabla f(x)$ and $Hf(x)$, respectively. We will use $Df(x)$ and $D^2 f(x)$ for the gradient and Hessian of $f(x)$, respectively, unless otherwise specified.

Vector Valued Functions

A function that takes values in a multidimensional space is called a vector valued function. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a vector valued function given by

$$F(x) \;=\; \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

where $x = (x_1, x_2, \ldots, x_n)$.

Definition 1.5. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ given by $F(x) = (f_j(x))_{j=1:m}$, and assume that the functions $f_j(x)$, $j = 1 : m$, are differentiable with respect to all variables x_i, $i = 1 : n$. The gradient $DF(x)$ of the function $F(x)$ is the following matrix of size $m \times n$:*

$$DF(x) \;=\; \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \ldots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}. \tag{1.38}$$

If $F : \mathbb{R}^n \to \mathbb{R}^n$, then the gradient $DF(x)$ is a square matrix of size n.

The j-th row of the gradient matrix $DF(x)$ is equal to the gradient $Df_j(x)$ of the function $f_j(x)$, $j = 1 : m$; cf. (1.36) and (1.38). Therefore,

$$DF(x) \;=\; \begin{pmatrix} Df_1(x) \\ Df_2(x) \\ \vdots \\ Df_m(x) \end{pmatrix}.$$

1.6.1 Functions of two variables

Functions of two variables are the simplest example of multivariable functions. To clarify the definitions for partial derivatives and for the gradient and the Hessian of multivariable functions given in section 1.6, we present them again for both scalar and vector valued functions of two variables.

Scalar Valued Functions

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables denoted by x and y. The partial derivatives of the function $f(x,y)$ with respect to the variables x and y are denoted by $\frac{\partial f}{\partial x}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$, respectively, and defined as follows:

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}; \\ \frac{\partial f}{\partial y}(x,y) &= \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}. \end{aligned}$$

The gradient of $f(x,y)$ is

$$Df(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{pmatrix}. \tag{1.39}$$

The Hessian of $f(x,y)$ is

$$D^2 f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial y \partial x}(x,y) \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix}. \tag{1.40}$$

Vector Valued Functions

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$F(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix}.$$

The gradient of $F(x,y)$ is

$$DF(x,y) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(x,y) & \frac{\partial f_1}{\partial y}(x,y) \\ \frac{\partial f_2}{\partial x}(x,y) & \frac{\partial f_2}{\partial y}(x,y) \end{pmatrix}. \tag{1.41}$$

Example: Let $f(x,y) = x^2y^3 + e^{2x+xy-1} - (x^3+3y^2)^2$. Evaluate the gradient and the Hessian of $f(x,y)$ at the point $(0,0)$.

Answer: By direct computation, we find that

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= 2xy^3 + (2+y)e^{2x+xy-1} - 6x^2(x^3+3y^2); \\ \frac{\partial f}{\partial y}(x,y) &= 3x^2y^2 + xe^{2x+xy-1} - 12y(x^3+3y^2); \\ \frac{\partial^2 f}{\partial x^2}(x,y) &= 2y^3 + (2+y)^2 e^{2x+xy-1} - 30x^4 - 36xy^2; \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) &= \frac{\partial}{\partial x}\left(3x^2y^2 + xe^{2x+xy-1} - 12y(x^3+3y^2)\right) \\ &= 6xy^2 + (1+2x+xy)e^{2x+xy-1} - 36x^2y; \\ \frac{\partial^2 f}{\partial y \partial x}(x,y) &= \frac{\partial}{\partial y}\left(2xy^3 + (2+y)e^{2x+xy-1} - 6x^2(x^3+3y^2)\right) \\ &= 6xy^2 + (1+2x+xy)e^{2x+xy-1} - 36x^2y; \\ \frac{\partial^2 f}{\partial y^2}(x,y) &= 6x^2y + x^2e^{2x+xy-1} - 12x^3 - 108y^2. \end{aligned}$$

Note that $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$, as stated by Theorem 1.9, since the function $f(x,y)$ is infinitely many times differentiable.

From (1.39) and (1.40), we find that

$$\begin{aligned} Df(0,0) &= \begin{pmatrix} \frac{\partial f}{\partial x}(0,0) & \frac{\partial f}{\partial y}(0,0) \end{pmatrix} = \begin{pmatrix} \frac{2}{e} & 0 \end{pmatrix}; \\ D^2 f(0,0) &= \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(0,0) & \frac{\partial^2 f}{\partial y \partial x}(0,0) \\ \frac{\partial^2 f}{\partial x \partial y}(0,0) & \frac{\partial^2 f}{\partial y^2}(0,0) \end{pmatrix} = \begin{pmatrix} \frac{4}{e} & \frac{1}{e} \\ \frac{1}{e} & 0 \end{pmatrix}. \quad \square \end{aligned}$$

FINANCIAL APPLICATIONS

Plain vanilla European call and put options.

The concept of arbitrage–free pricing.

Pricing European plain vanilla options if the underlying asset is worthless.

Put–Call parity for European options.

Forward and Futures contracts.

1.7 Plain vanilla European Call and Put options

A **Call Option** on an underlying asset (e.g., on one share of a stock, for an equity option[4]) is a contract between two parties which gives the buyer of the option the right, but not the obligation, to **buy** from the seller of the option one unit of the asset (e.g., one share of the stock) at a predetermined time T in the future, called the maturity of the option, for a predetermined price K, called the strike of the option. For this right, the buyer of the option pays $C(t)$ at time $t < T$ to the seller of the option.

A **Put Option** on an underlying asset is a contract between two parties which gives the buyer of the option the right, but not the obligation, to **sell** to the seller of the option one unit of the asset at a predetermined time T in the future, called the maturity of the option, for a predetermined price K, called the strike of the option. For this right, the buyer of the option pays $P(t)$ at time $t < T$ to the seller of the option.

The options described above are plain vanilla European options. An American option can be exercised at any time prior to maturity.

In an option contract, two parties exist: the buyer of the option and the seller of the option. We also say that the buyer of the option is long the option (or has a long position in the option) and that the seller of the option is short the option (or has a short position in the option).

Let $S(t)$ and $S(T)$ be the price of the underlying asset at time t and at maturity T, respectively.

At time t, a call option is in the money (ITM), at the money (ATM), or out of the money (OTM), depending on whether $S(t) > K$, $S(t) = K$, or $S(t) < K$, respectively, A put option is in the money, at the money, or out of the money at time t if $S(t) < K$, $S(t) = K$, or $S(t) > K$, respectively.

At maturity T, a call option expires in the money (ITM), at the money (ATM), or out of the money (OTM), depending on whether $S(T) > K$, $S(T) = K$, or $S(T) < K$, respectively. A put option expires in the money, at the money, or out of the money, if $S(T) < K$, $S(T) = K$, or $S(T) > K$, respectively.

The payoff of a call option at maturity is

$$C(T) \;=\; \max(S(T) - K, 0) \;=\; \begin{cases} S(T) - K, & \text{if } S(T) > K; \\ 0, & \text{if } S(T) \leq K. \end{cases}$$

The payoff of a put option at maturity is

$$P(T) \;=\; \max(K - S(T), 0) \;=\; \begin{cases} 0, & \text{if } S(T) \geq K; \\ K - S(T), & \text{if } S(T) < K. \end{cases}$$

[4] The underlying asset for equity options is usually 100 shares, not one share. For clarity and simplicity reasons, we will be consistent throughout the book in our assumption that options are written on just one unit of the underlying asset.

1.8 Arbitrage–free pricing

An arbitrage opportunity is an investment opportunity that is guaranteed to earn money without any risk involved. While such arbitrage opportunities exist in the markets, many of them are of little practical value. Trading costs, lack of liquidity, the bid–ask spread, constant moves of the market that tend to quickly eliminate any arbitrage opportunity, and the impossibility of executing large enough trades without moving the markets make it very difficult to capitalize on arbitrage opportunities.

In an arbitrage–free market, we can infer relationships between the prices of various securities, based on the following principle:

Theorem 1.10. ***(The (Generalized) Law of One Price.)*** *If two portfolios are guaranteed to have the same value at a future time $\tau > t$ regardless of the state of the market at time τ, then they must have the same value at time t. If one portfolio is guaranteed to be more valuable (or less valuable) than another portfolio at a future time $\tau > t$ regardless of the state of the market at time τ, then that portfolio is more valuable (or less valuable, respectively) than the other one at time $t < \tau$ as well:*

If there exists $\tau > t$ such that $V_1(\tau) = V_2(\tau)$ (or $V_1(\tau) > V_2(\tau)$, or $V_1(\tau) < V_2(\tau)$, respectively) for any state of the market at time τ, then $V_1(t) = V_2(t)$ (or $V_1(t) > V_2(t)$, or $V_1(t) < V_2(t)$, respectively).

Corollary 1.1. *If the value of a portfolio of securities is guaranteed to be equal to 0 at a future time $\tau > t$ regardless of the state of the market at time τ, then the value of the portfolio at time t must have been 0 as well:*

If there exists $\tau > t$ such that $V(\tau) = 0$ for any state of the market at time τ, then $V(t) = 0$.

An important consequence of the law of one price is the fact that, if the value of a portfolio at time T in the future is independent of the state of the market at that time, then the value of the portfolio in the present is the risk–neutral discounted present value of the portfolio at time T.

Before we state this result formally, we must clarify the meaning of "risk–neutral discounted present value". This refers to the time value of money: cash can be deposited at time t_1 to be returned at time t_2 ($t_2 > t_1$), with interest. The interest rate depends on many factors, one of them being the probability of default of the party receiving the cash deposit. If this probability is zero, or close to zero (the US Treasury is considered virtually impossible to default – more money can be printed to pay back debt, for example), then the return is considered risk–free. Interest can be compounded in many different ways, e.g., annual, semi-annual, continuous. Unless otherwise specified, throughout this book, interest is assumed to be compounded continuously.

For continuously compounded interest, the value $B(t_2)$ at time $t_2 > t_1$ of $B(t_1)$ cash at time t_1 is

$$B(t_2) \;=\; e^{r(t_2-t_1)}\, B(t_1), \tag{1.42}$$

where r is the risk free rate between time t_1 and t_2. The value $B(t_1)$ at time $t_1 < t_2$ of $B(t_2)$ cash at time t_2 is

$$B(t_1) \;=\; e^{-r(t_2-t_1)}\, B(t_2). \tag{1.43}$$

More details on interest rates are given in section 2.6. Formulas (1.42) and (1.43) are the same as formulas (2.46) and (2.48) from section 2.6.

Lemma 1.7. *If the value $V(T)$ of a portfolio at time T in the future is independent of the state of the market at time T, then*

$$V(t) \;=\; V(T)\, e^{-r(T-t)}, \tag{1.44}$$

where $t < T$ and r is the constant risk free rate.

Proof. For clarity purposes, let $z = V(T)$ be the value of the portfolio at time T. Consider a portfolio made of $V_2(t) = ze^{-r(T-t)}$ cash at time t. The value $V_2(T)$ of this portfolio at time T is

$$V_2(T) \;=\; e^{r(T-t)}\, V_2(t) \;=\; e^{r(T-t)}\left(ze^{-r(T-t)}\right) \;=\; z;$$

cf. (1.42) for $t_1 = t$, $t_2 = T$, and $B(t) = V_2(t)$.

Thus, $V_2(T) = V(T) = z$, and, from Theorem 1.10, we conclude that $V_2(t) = V(t)$. Therefore, $V(t) = V_2(t) = ze^{-r(T-t)} = V(T)\, e^{-r(T-t)}$, which is what we wanted to prove. □

Example: How much are plain vanilla European options worth if the value of the underlying asset is 0?

Answer: If, at time t, the underlying asset becomes worthless, i.e., if $S(t) = 0$, then the price of the asset will never be above 0 again. Otherwise, an arbitrage opportunity arises: buy the asset at no cost at time t, and sell it for risk–free profit as soon as its value is above 0.

In particular, at maturity, the spot price will be zero, i.e., $S(T) = 0$. Then, at maturity, the call option will be worthless, while the put option will always be exercised for a premium of K, i.e.,

$$\begin{aligned} C(T) &= \max(S(T)-K,0) \;=\; 0 \\ P(T) &= \max(K-S(T),0) \;=\; K. \end{aligned}$$

From Lemma 1.7, we conclude that

$$\begin{aligned} C(t) &= 0; \\ P(t) &= Ke^{-r(T-t)}, \end{aligned} \tag{1.45}$$

where r is the constant risk free rate. □

1.9 The Put–Call parity for European options

Let $C(t)$ and $P(t)$ be the values at time t of a European call and put option, respectively, with maturity T and strike K, on the same non–dividend paying asset with spot price $S(t)$. The Put–Call parity states that

$$P(t) \;+\; S(t) \;-\; C(t) \;=\; Ke^{-r(T-t)}. \tag{1.46}$$

If the underlying asset pays dividends continuously at the rate q, the Put–Call parity has the form

$$P(t) \;+\; S(t)e^{-q(T-t)} \;-\; C(t) \;=\; Ke^{-r(T-t)}. \tag{1.47}$$

We prove (1.46) here using the law of one price.

Consider a portfolio made of the following assets:

- long 1 put option;
- long 1 share;
- short 1 call option.

The value[5] of the portfolio at time t is

$$V_{portfolio}(t) \;=\; P(t) + S(t) - C(t). \tag{1.48}$$

It is easy to see that

$$V_{portfolio}(T) \;=\; P(T) + S(T) - C(T) \;=\; K, \tag{1.49}$$

regardless of the value $S(T)$ of the underlying asset at the maturity of the option, e.g., by analyzing what happens if $S(T) < K$ or if $S(T) \geq K$:

From Lemma 1.7 and (1.49), we obtain that

$$V_{portfolio}(t) \;=\; V_{portfolio}(T)e^{-r(T-t)} \;=\; Ke^{-r(T-t)}. \tag{1.50}$$

[5]It is important to clarify that the value of a portfolio is equal to the cash amount generated if the portfolio is liquidated, and not to the cash amount needed to set up the portfolio. For example, if you own a portfolio consisting of long one call option with price C, the value of the portfolio is $+C$, since this is how much would be obtained by selling the call option, and not $-C$, which is the amount needed to buy the call and set up the portfolio.

	$P(T)$	$C(T)$	$P(T) + S(T) - C(T)$
$S(T) < K$	$K - S(T)$	0	$(K - S(T)) + S(T) - 0 = K$
$S(T) \geq K$	0	$S(T) - K$	$0 + S(T) - (S(T) - K) = K$

The Put–Call parity formula (1.46) follows from (1.48) and (1.50):

$$P(t) \ + \ S(t) \ - \ C(t) \ = \ Ke^{-r(T-t)}.$$

1.10 Forward and Futures contracts

A forward contract is an agreement between two parties: one party (the long position) agrees to buy the underlying asset from the other party (the short position) at a specified time in the future and for a specified price, called the forward price. The forward price is chosen such that the forward contract has value zero at the time when the forward contract is entered into. (It is helpful to think of the forward price as the contractual forward price which is set at the inception of the forward contract as the delivery price.) Note that the forward price is *not* the price of the forward contract.

We will show that the contractual forward price F of a forward contract with maturity T and struck at time 0 on a non–dividend–paying underlying asset with spot price $S(0)$ is

$$F \ = \ S(0)e^{rT}. \tag{1.51}$$

Here, the interest rate is assumed to be constant and equal to r over the life of the forward contract, i.e., between times 0 and T.

If the underlying asset pays dividends continuously at the rate q, the forward price is

$$F \ = \ S(0)e^{(r-q)T}. \tag{1.52}$$

A futures contract has a similar structure as a forward contract, but it requires the delivery of the underlying asset for the futures price. (Forward contracts can be settled in cash at maturity, without the delivery of a physical asset.) The forward and futures prices are, in theory, the same, if the risk–free interest rates are constant or deterministic, i.e., just functions of time. Several major differences exist between the ways forward and futures contracts are structured, settled, and traded:

- Futures contracts trade on an exchange and have standard features, while forward contracts are over-the-counter instruments;

- Futures are marked to market and settled in a margin account on a daily basis, while forward contracts are settled in cash at maturity;
- Futures have a range of delivery dates, while forward contracts have a specified delivery date;
- Futures carry almost no credit risk, since they are settled daily, while entering into a forward contract carries some credit risk[6].

To derive formula (1.51), consider a forward contract written at time 0 and requiring to buy one unit of the underlying asset at time T for the forward price F. The value at time 0 of a portfolio consisting of
• long 1 forward contract;
• short 1 unit of the underlying asset
is $V(0) = -S(0)$, since the value of the forward at inception is equal to 0.

At maturity, the forward contract is exercised. One unit of the underlying asset is bought for the contractual forward price F and the short position in the asset is closed. The portfolio will be all cash at time T, and its value will be $V(T) = -F$, regardless of what the price $S(T)$ of the underlying asset is at maturity. From Lemma 1.7, we obtain that

$$V(0) \ = \ e^{-r(T-t)}V(T),$$

where r is the constant risk free rate. Since $V(0) = -S(0)$ and $V(T) = -F$, we conclude that $F = S(0)e^{rT}$, and formula (1.51) is proven.

A similar argument can be used to value a forward contract that was struck at an earlier date. Consider a forward contract with delivery price K that will expire at time T. (One unit of the underlying asset will be bought at time T for the price K by the long position.) Let $F(t)$ be the value of the forward contract at time t. Consider a portfolio made of the following assets:
• long 1 forward contract;
• short 1 unit of the underlying asset;
The value of the portfolio at time t is $V(t) = F(t) - S(t)$. Note that $F(t)$, the value of the forward contract at time t, is no longer equal to 0, since the forward contract was not written at time t, but in the past.

At maturity T, the forward contract is exercised and the short position is closed: we pay K for one unit of the underlying asset and return it to the lender. The value $V(T)$ of the portfolio at time T is $V(T) = -K$. Since $V(T)$ is a cash amount independent of the price $S(T)$ of the underlying asset at maturity, we find from Lemma 1.7 that

$$V(t) \ = \ V(T)e^{-r(T-t)}.$$

[6]The value of a forward is 0 at inception, but changes over time. The credit risk comes from the risk of default of the party for whom the value of the forward contract is negative.

Since $V(t) = F(t) - S(t)$ and $V(T) = -K$, we conclude that the price at time t of a forward contract with delivery price K and maturity T is

$$F(t) = S(t) - Ke^{-r(T-t)}. \tag{1.53}$$

If the underlying asset pays dividends continuously at the rate q, the value of the forward contract is

$$F(t) = S(t)e^{-qT} - Ke^{-r(T-t)}.$$

It is interesting to note the connection between forward contracts and the Put–Call parity (1.46).

Being long a call and short a put with the same strike K is equivalent to being long a forward contract with delivery price K. To see this, note that the value $F(T)$ of a forward contract at delivery time T is $F(T) = S(T) - K$, since the amount K is paid for one unit of the underlying asset. The value at time T of a long call and short put position is

$$C(T) - P(T) = \max(S(T) - K, 0) - \max(K - S(T), 0) = S(T) - K = F(T).$$

Thus, $C(T) - P(T) = F(T)$ for any value $S(T)$ of the underlying asset at maturity. From Theorem 1.10 and (1.53), it follows that

$$C(t) - P(t) = F(t) = S(t) - Ke^{-r(T-t)},$$

which is the same as the Put–Call parity (1.46).

1.11 References

Most of the mathematical topics from this chapter, as well as from the rest of the book, appear in many calculus advanced books, such as Edwards [10] and Protter and Morrey [20], where they are presented at different levels of mathematical sophistication.

Two classical texts covering a wide range of financial products, from plain vanilla options to credit derivatives, are Hull [14] and Neftci [18]. Another book by Hull [13] provides details on futures trading, while Neftci [19] gives a practitioner's perspective on financial instruments. Another mathematical finance book is Joshi [16]. A personal view on quantitative finance from a leading practitioner and educator can be found in the introductory text Wilmott [34], as well as in the comprehensive three volume monograph Wilmott [33].

1.12 Exercises

1. Use the integration by parts to compute $\int \ln(x)\, dx$.

2. Compute $\int \frac{1}{x\ln(x)}\, dx$ by using the substitution $u = \ln(x)$.

3. Show that $(\tan x)' = 1/(\cos x)^2$ and

$$\int \frac{1}{1+x^2}\, dx = \arctan(x) + C.$$

Note: The antiderivative of a rational function is often computed using the substitution $x = \tan\left(\frac{z}{2}\right)$.

4. Use l'Hôpital's rule to show that the following two Taylor approximations hold when x is close to 0:

$$\sqrt{1+x} \approx 1 + \frac{x}{2};$$
$$e^x \approx 1 + x + \frac{x^2}{2}.$$

In other words, show that the following limits exist and are constant:

$$\lim_{x\to 0} \frac{\sqrt{1+x} - \left(1 + \frac{x}{2}\right)}{x^2} \quad \text{and} \quad \lim_{x\to 0} \frac{e^x - \left(1 + x + \frac{x^2}{2}\right)}{x^3}.$$

5. Use the definition (1.32) of e, i.e., $e = \lim_{x\to\infty}\left(1 + \frac{1}{x}\right)^x$, to show that

$$\frac{1}{e} = \lim_{x\to\infty}\left(1 - \frac{1}{x}\right)^x.$$

Hint: Use the fact that

$$\frac{1}{1 + \frac{1}{x}} = \frac{x}{x+1} = 1 - \frac{1}{x+1}.$$

6. Let K, T, σ and r be positive constants, and define the function $g : \mathbb{R} \to \mathbb{R}$ as

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_0^{b(x)} e^{-\frac{y^2}{2}}\, dy,$$

where $b(x) = \left(\ln\left(\frac{x}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T\right) / \left(\sigma\sqrt{T}\right)$. Compute $g'(x)$.

Note: This function is related to the Delta of a plain vanilla Call option; see Section 3.5 for more details.

7. Let $f(x)$ be a continuous function. Show that

$$\lim_{h\to 0} \frac{1}{2h} \int_{a-h}^{a+h} f(x)\, dx = f(a), \quad \forall\ a \in \mathbb{R}.$$

Note: Let $F(x) = \int f(x)\, dx$. The central finite difference approximation (6.7) of $F'(a)$ is

$$F'(a) = \frac{F(a+h) - F(a-h)}{2h} + O\left(h^2\right), \qquad (1.54)$$

as $h \to 0$ (if $F^{(3)}(x) = f''(x)$ is continuous). Since $F'(a) = f(a)$, formula (1.54) can be written as

$$f(a) = \frac{1}{2h} \int_{a-h}^{a+h} f(x)\, dx + O(h^2).$$

8. Let $f : \mathbb{R} \to \mathbb{R}$ given by $f(y) = \sum_{i=1}^{n} c_i e^{-yt_i}$, where c_i and t_i, $i = 1 : n$, are positive constants. Compute $f'(y)$ and $f''(y)$.

Note: The function $f(y)$ represents the price of a bond with cash flows c_i paid at time t_i as a function of the yield y of the bond. When scaled appropriately, the derivative of $f(y)$ with respect to y give the duration and convexity of the bond; see Section 2.7 for more details.

9. Let $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x) = 2x_1^2 - x_1x_2 + 3x_2x_3 - x_3^2$, where $x = (x_1, x_2, x_3)$.

(i) Compute the gradient and Hessian of the function $f(x)$ at the point $a = (1, -1, 0)$, i.e., compute $Df(1, -1, 0)$ and $D^2 f(1, -1, 0)$.

(ii) Show that

$$f(x) = f(a) + Df(a)\,(x - a) + \frac{1}{2}\,(x - a)^t\, D^2 f(a)\,(x - a). \qquad (1.55)$$

Here, x, a, and $x - a$ are 3×1 column vectors, i.e.,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}; \quad a = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}; \quad x - a = \begin{pmatrix} x_1 - 1 \\ x_2 + 1 \\ x_3 \end{pmatrix}.$$

Note: Formula (1.55) is the quadratic Taylor approximation of $f(x)$ around the point a; cf. (5.32). Since $f(x)$ is a second order polynomial, the quadratic Taylor approximation of $f(x)$ is exact.

10. Let

$$u(x,t) = \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{x^2}{4t}}, \quad \text{for } t > 0,\ x \in \mathbb{R}.$$

Compute $\frac{\partial u}{\partial t}$ and $\frac{\partial^2 u}{\partial x^2}$, and show that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}.$$

Note: This exercise shows that the function $u(x,t)$ is a solution of the heat equation. In fact, $u(x,t)$ is the fundamental solution of the heat equation, and is used in the PDE derivation of the Black–Scholes formula for pricing European plain vanilla options.

Also, note that $u(x,t)$ is the same as the density function of a normal variable with mean 0 and variance $2t$; cf. (3.48) for $\mu = 0$ and $\sigma^2 = 2t$.

11. Consider a portfolio with the following positions:

- long one call option with strike $K_1 = 30$;
- short two call options with strike $K_2 = 35$;
- long one call option with strike $K_3 = 40$.

All options are on the same underlying asset and have maturity T. Draw the payoff diagram at maturity of the portfolio, i.e., plot the value of the portfolio $V(T)$ at maturity as a function of $S(T)$, the price of the underlying asset at time T.

Note: This is a butterfly spread. A trader takes a long position in a butterfly spread if the price of the underlying asset at maturity is expected to be in the $K_1 \leq S(T) \leq K_3$ range.

12. Draw the payoff diagram at maturity of a bull spread with a long position in a call with strike 30 and short a call with strike 35, and of a bear spread with long a put of strike 20 and short a put of strike 15.

13. Which of the following two portfolios would you rather hold:

- Portfolio 1: Long one call option with strike $K = X - 5$ and long one call option with strike $K = X + 5$;

- Portfolio 2: Long two call options with strike $K = X$?

(All options are on the same asset and have the same maturity.)

14. Call options with strikes 100, 120, and 130 on the same underlying asset and with the same maturity are trading for 8, 5, and 3, respectively (there is no bid–ask spread). Is there an arbitrage opportunity present? If yes, how can you make a riskless profit?

15. A stock with spot price 40 pays dividends continuously at a rate of 3%. The four months at-the-money put and call options on this asset are trading at \$2 and \$4, respectively. The risk-free rate is constant and equal to 5% for all times. Show that the Put-Call parity is not satisfied and explain how would you take advantage of this arbitrage opportunity.

16. The bid and ask prices for a six months European call option with strike 40 on a non–dividend–paying stock with spot price 42 are \$5 and \$5.5, respectively. The bid and ask prices for a six months European put option with strike 40 on the same underlying asset are \$2.75 and \$3.25, respectively. Assume that the risk free rate is equal to 0. Is there an arbitrage opportunity present?

17. You expect that an asset with spot price \$35 will trade in the \$40–\$45 range in one year. One year at–the–money calls on the asset can be bought for \$4. To act on the expected stock price appreciation, you decide to either buy the asset, or to buy ATM calls. Which strategy is better, depending on where the asset price will be in a year?

18. The risk free rate is 8% compounded continuously and the dividend yield of a stock index is 3%. The index is at 12,000 and the futures price of a contract deliverable in three months is 12,100. Is there an arbitrage opportunity, and how do you take advantage of it?

Chapter 2

Improper integrals. Numerical integration. Interest rates. Bonds.

Double integrals. Switching the order of integration.

Convergence and evaluation of improper integrals.

Differentiating improper integrals with respect to the integration limits.

Numerical methods for computing definite integrals: the Midpoint, Trapezoidal, and Simpson's rules. Convergence and numerical implementation.

2.1 Double integrals

Let $D \subset \mathbb{R}^2$ be a bounded region and let $f : D \to \mathbb{R}$ be a continuous function. The double integral of f over D, denoted by

$$\int\int_D f,$$

represents the volume of the three dimensional body between the domain D in the two dimensional plane and the graph of the function $f(x, y)$.

The double integral of f over D can be computed first with respect to the variable x, and then with respect to the variable y, when it is denoted as

$$\int\int_D f(x,y)\ dx\ dy,$$

or can be computed first with respect to the variable y, and then with respect to the variable x, when it is denoted as

$$\int\int_D f(x,y)\ dy\ dx.$$

We will define these integrals properly in (2.2) and (2.3), and specify an instance when the two integrals have the same value and are equal to the double integral of f over D in Theorem 2.1.

For simplicity, assume that the domain D is bounded and convex, i.e., for any two points x_1 and x_2 in D, all the points on the segment joining x_1 and x_2 are in D as well. Also, assume that there exist two continuous functions $f_1(x)$ and $f_2(x)$ such that D can be described as follows:

$$D = \{(x, y) \mid a \leq x \leq b \text{ and } f_1(x) \leq y \leq f_2(x)\}. \qquad (2.1)$$

The functions $f_1(x)$ and $f_2(x)$ are well defined by (2.1) since the domain D is bounded and convex. Then, by definition,

$$\int\int_D f(x,y)\, dy\, dx = \int_a^b \left(\int_{f_1(x)}^{f_2(x)} f(x,y)\, dy\right) dx. \qquad (2.2)$$

If there exist two continuous functions $g_1(y)$ and $g_2(y)$ such that $D = \{(x,y) \mid c \leq y \leq d \text{ and } g_1(y) \leq x \leq g_2(y)\}$, then, by definition,

$$\int\int_D f(x,y)\, dx\, dy = \int_c^d \left(\int_{g_1(y)}^{g_2(y)} f(x,y)\, dx\right) dy. \qquad (2.3)$$

Theorem 2.1. ***(Fubini's Theorem.)*** *With the notations above, if the function $f(x,y)$ is continuous, then the integrals (2.2) and (2.3) are equal to each other and to the double integral of $f(x,y)$ over D, i.e., the order of integration does not matter:*

$$\int\int_D f = \int\int_D f(x,y)\, dx\, dy = \int\int_D f(x,y)\, dy\, dx.$$

For example, if $D = [a,b] \times [c,d]$ is a rectangle and if $f(x,y)$ is a continuous function, then

$$\int\int_D f = \int_a^b \int_c^d f(x,y)\, dy\, dx = \int_c^d \int_a^b f(x,y)\, dx\, dy.$$

Example: Let $D = [1,3] \times [2,5]$ and $f(x,y) = 2y - 3x$. Compute $\int\int_D f$.

Answer: Since $f(x,y)$ is continuous, it follows from Theorem 2.1 that it is enough to compute either $\int\int_D f(x,y) dx dy$ or $\int\int_D f(x,y) dy dx$. Thus,

$$\int\int_D f = \int_2^5 \left(\int_1^3 (2y - 3x) dx\right) dy = \int_2^5 (4y - 12) dy = 6.$$
$$\int\int_D f = \int_1^3 \left(\int_2^5 (2y - 3x) dy\right) dx = \int_1^3 (21 - 9x) dx = 6. \quad \square$$

Example: Assume D is a disk of radius 2 centered at 0 and let $f(x,y) = 1 - x^2 - y^2$. Compute $\int\int_D f$.

Answer: Note that $D = \{(x,y) : x^2 + y^2 \leq 4\}$. The points on the boundary of D (i.e., on the circle of center 0 and radius 2) satisfy $x^2 + y^2 = 4$. Solving for y we find that $y^2 = 4 - x^2$ and therefore $y = \pm\sqrt{4 - x^2}$. This corresponds to $a = -2$, $b = 2$, $f_1(x) = -\sqrt{4 - x^2}$ and $f_2(x) = \sqrt{4 - x^2}$ in (2.1), i.e.,

$$D = \{(x,y) : -2 \leq x \leq 2 \text{ and } -\sqrt{4 - x^2} \leq y \leq \sqrt{4 - x^2}\}.$$

Then, from Theorem 2.1, it follows that

$$\begin{aligned}
\int\int_D f &= \int\int_D f(x,y)\, dy\, dx = \int_{-2}^{2} \left(\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} 1 - x^2 - y^2\, dy\right) dx \\
&= \int_{-2}^{2} \left((1 - x^2) y - \frac{y^3}{3}\right)\Bigg|_{y=-\sqrt{4-x^2}}^{y=\sqrt{4-x^2}} dx \\
&= \int_{-2}^{2} 2(1 - x^2)\sqrt{4 - x^2} - \frac{2(\sqrt{4 - x^2})^3}{3}\, dx \\
&= -\frac{2}{3}\int_{-2}^{2} (1 + 2x^2)\sqrt{4 - x^2}\, dx = -\frac{4}{3}\int_0^2 (1 + 2x^2)\sqrt{4 - x^2}\, dx;
\end{aligned}$$

the last equality follows from Lemma 0.1, since $(1 - 2x^2)\sqrt{4 - x^2}$ is an even function. After making the change of variables $x = 2\sin(t)$, for $t \in [0, \frac{\pi}{2}]$, we find out that

$$\int\int_D f = -4\pi.$$

We note that this integral will be computed in a more straightforward fashion using polar coordinates; see section 7.2.1. $\square$

Let $D = [a,b] \times [c,d]$ be a rectangle and let $f(x,y)$ be a continuous function with separated variables, i.e., $f(x,y) = g(x)h(y)$ with $g(x)$ and $h(y)$ continuous. Then the two dimensional integral of $f(x,y)$ over D is the product of two one dimensional integrals:

$$\begin{aligned}
\int\int_D f(x,y)\, dx\, dy &= \int_c^d \int_a^b g(x)h(y)\, dx\, dy = \int_c^d h(y) \left(\int_a^b g(x) dx\right) dy \\
&= \left(\int_a^b g(x)\, dx\right) \left(\int_c^d h(y)\, dy\right). \qquad (2.4)
\end{aligned}$$

Note that this does not happen if $f(x,y)$ is separable but D is not a rectangle. Let $D = \{(x,y) : c \leq y \leq d,\ g_1(y) \leq x \leq g_2(y)\}$ and assume that $f(x,y) = g(x)h(y)$. Then,

$$\int\int_D f(x,y)\, dx\, dy = \int_c^d \int_{g_1(y)}^{g_2(y)} g(x)h(y)\, dx\, dy$$

$$= \int_c^d h(y) \left(\int_{g_1(y)}^{g_2(y)} g(x)\, dx \right) dy.$$

The integral $\int_{g_1(y)}^{g_2(y)} g(x)\, dx$ is, in general, a function of the variable y. Therefore, we cannot take the term $\int_{g_1(y)}^{g_2(y)} g(x)\, dx$ outside the sign of integration with respect to y, as we did in (2.4).

2.2 Improper integrals

We consider three types of improper integrals:

• Integrate the function $f(x)$ over an infinite interval of the form $[a, \infty)$ or $(-\infty, b]$. The integral $\int_a^\infty f(x)\, dx$ exists if and only if the limit as $t \to \infty$ of the definite integral of $f(x)$ between a and t exists and is finite. The integral $\int_{-\infty}^b f(x)\, dx$ exists if and only if the limit as $t \to -\infty$ of the definite integral of $f(x)$ between t and b exists and is finite. Then,

$$\int_a^\infty f(x)\, dx = \lim_{t\to\infty} \int_a^t f(x)\, dx; \tag{2.5}$$

$$\int_{-\infty}^b f(x)\, dx = \lim_{t\to-\infty} \int_t^b f(x)\, dx. \tag{2.6}$$

Adding and subtracting improper integrals of this type follows rules similar to those for definite integrals:

Lemma 2.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be an integrable function over the interval $[a, \infty)$. If $b > a$, then $f(x)$ is also integrable over the interval $[b, \infty)$ and*

$$\int_a^\infty f(x)\, dx - \int_b^\infty f(x)\, dx = \int_a^b f(x)\, dx.$$

Let $f(x)$ be an integrable function over the interval $(-\infty, b]$. If $a < b$, then $f(x)$ is also integrable over the interval $(-\infty, a]$ and

$$\int_{-\infty}^b f(x)\, dx - \int_{-\infty}^a f(x)\, dx = \int_a^b f(x)\, dx.$$

• Integrate the function $f(x)$ over an interval $[a, b]$ where $f(x)$ is unbounded as x approaches the end points a and/or b. For example, if the limit as $x \searrow a$

of $f(x)$ is infinite, then $\int_a^b f(x)\, dx$ exists if and only if the limit as $t \searrow a$ of the definite integral of $f(x)$ between t and b exists and is finite, i.e.,

$$\int_a^b f(x)\, dx = \lim_{t\searrow a} \int_t^b f(x)\, dx.$$

• Integrate the function $f(x)$ on the entire real axis, i.e., on $(-\infty, \infty)$. The integral $\int_{-\infty}^\infty f(x)\, dx$ exists if and only if there exists a real number a such that both $\int_{-\infty}^a f(x)\, dx$ and $\int_a^\infty f(x)\, dx$ exist. Then,

$$\int_{-\infty}^\infty f(x)\, dx = \int_{-\infty}^a f(x)\, dx + \int_a^\infty f(x)\, dx \tag{2.7}$$

$$= \lim_{t_1\to-\infty} \int_{t_1}^a f(x)\, dx + \lim_{t_2\to\infty} \int_a^{t_2} f(x)\, dx. \tag{2.8}$$

Note that it is incorrect to use, instead of (2.8), the following definition for the integral of $f(x)$ over the real axis $(-\infty, \infty)$:

$$\int_{-\infty}^\infty f(x)\, dx = \lim_{t\to\infty} \int_{-t}^t f(x)\, dx. \tag{2.9}$$

For a simple example, it is easy to see that

$$\int_{-t}^t x\, dx = 0, \quad \forall\, t > 0,$$

while $\int_{-\infty}^0 x\, dx = -\infty$ and $\int_0^\infty x\, dx = \infty$, for any $a \in \mathbb{R}$. Therefore, according to definition (2.7), $\int_{-\infty}^\infty x\, dx$ cannot be defined.

However, if we know that the function $f(x)$ in integrable over the entire real axis, then we can use formula (2.9) to evaluate it:

Lemma 2.2. *If the improper integral $\int_{-\infty}^\infty f(x)dx$ exists, then*

$$\int_{-\infty}^\infty f(x)\, dx = \lim_{t\to\infty} \int_{-t}^t f(x)\, dx. \tag{2.10}$$

Example: Show that, for any $\alpha > 0$,

$$\int_0^\infty x^\alpha e^{-x^2}\, dx \tag{2.11}$$

exists and is finite, and conclude that

$$\int_{-\infty}^{\infty} x^{2k} e^{-x^2}\, dx$$

exists and is finite for any positive integer k. (Integrals of this type need to be evaluated when computing the expected value and variance of the standard normal variable; see (3.33) and (3.34).)

Answer: Let $h : [0, \infty) \to \mathbb{R}$ be given by

$$h(t) \;=\; \int_0^t x^\alpha e^{-x^2}\, dx.$$

By definition (2.5), to prove that $\int_0^\infty x^\alpha e^{-x^2}\, dx$ exists and is finite, it is enough to show that

$$\lim_{t\to\infty} \int_0^t x^\alpha e^{-x^2}\, dx \;=\; \lim_{t\to\infty} h(t)$$

exists and is finite. (The function $x^\alpha e^{-x^2}$ is bounded as $x \searrow 0$ since $\alpha > 0$.)

Since exponential functions increase much faster than power functions as x goes to infinity, it follows that

$$\lim_{x\to\infty} x^{\alpha+2}\, e^{-x^2} \;=\; 0.$$

Then, there exists a constant $M > 0$ such that

$$x^\alpha e^{-x^2} \;<\; \frac{1}{x^2}, \quad \forall\, x \geq M.$$

For any $t > M$ we obtain that

$$\begin{aligned} h(t) \;&=\; \int_0^M x^\alpha e^{-x^2}\, dx \;+\; \int_M^t x^\alpha e^{-x^2}\, dx \\ &<\; h(M) \;+\; \int_M^t \frac{1}{x^2}\, dx \;=\; h(M) \;+\; \left(-\frac{1}{t} + \frac{1}{M}\right) \\ &<\; h(M) \;+\; \frac{1}{M}. \end{aligned}$$

We conclude that the function $h(t)$ is bounded as $t \to \infty$. Since $h(t)$ is also increasing, it follows that $\lim_{t\to\infty} h(t)$ exists and is finite, and therefore that $\int_0^\infty x^\alpha e^{-x^2}\, dx$ exists and is finite.

Note that $\int_0^\infty x^{2k} e^{-x^2}\, dx$ exists and is finite; cf. (2.11) for $\alpha = 2k$. Then, from (5) of Lemma 0.1, it follows that

$$\int_{-\infty}^{\infty} x^{2k} e^{-x^2}\, dx \;=\; 2\int_0^\infty x^{2k} e^{-x^2}\, dx,$$

and therefore that $\int_{-\infty}^{\infty} x^{2k} e^{-x^2}\, dx$ exists and is finite.

Examples: Compute

$$\int_1^\infty e^{-x}\, dx; \quad \int_1^\infty e^{x}\, dx; \quad \int_0^1 \frac{1}{\sqrt{x}}\, dx; \quad \int_{-\infty}^0 \frac{1}{(2x-1)^3}\, dx; \quad \int_{-\infty}^\infty e^{-|x|}\, dx.$$

Answers:

$$\begin{aligned} \int_1^\infty e^{-x}\, dx \;&=\; \lim_{t\to\infty} \int_1^t e^{-x}\, dx \;=\; \lim_{t\to\infty} -e^{-t} + e^{-1} \;=\; \frac{1}{e}; \\ \int_1^\infty e^{x}\, dx \;&=\; \lim_{t\to\infty} \int_1^t e^{x}\, dx \;=\; \lim_{t\to\infty} e^{x} - e^{1} \;=\; \infty; \\ \int_0^1 \frac{1}{\sqrt{x}}\, dx \;&=\; \lim_{t\searrow 0} \int_t^1 \frac{1}{\sqrt{x}}\, dx \;=\; \lim_{t\searrow 0} 2 - 2\sqrt{t} \;=\; 2; \\ \int_{-\infty}^0 \frac{1}{(2x-1)^3}\, dx \;&=\; \lim_{t\to-\infty} \int_t^0 \frac{1}{(2x-1)^3}\, dx \\ &=\; \lim_{t\to-\infty} -\frac{1}{4(2x-1)^2}\Big|_t^0 \;=\; -\frac{1}{4}; \\ \int_{-\infty}^\infty e^{-|x|}\, dx \;&=\; \int_{-\infty}^0 e^{-|x|}\, dx \;+\; \int_0^\infty e^{-|x|}\, dx \\ &=\; \int_{-\infty}^0 e^{x}\, dx \;+\; \int_0^\infty e^{-x}\, dx \\ &=\; \lim_{y\to-\infty} \int_y^0 e^{x}\, dx \;+\; \lim_{z\to\infty} \int_0^z e^{-x}\, dx \\ &=\; \lim_{y\to-\infty} e^{x}\,\Big|_y^0 \;+\; \lim_{z\to\infty} \left(-e^{-x}\right)\,\Big|_0^z \\ &=\; \lim_{y\to-\infty} (1 - e^{y}) \;+\; \lim_{z\to\infty} \left(-e^{-z} + 1\right) \\ &=\; 2. \end{aligned}$$

2.3 Differentiating improper integrals with respect to the integration limits

This topic appears frequently in conjunction with the Black–Scholes pricing model for plain vanilla options, e.g., when exact formulas for the Greeks are computed; see Section 3.6 for more details.

Lemma 2.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that the improper integral $\int_{-\infty}^{\infty} f(x)\ dx$ exists. Let $g, h : \mathbb{R} \to \mathbb{R}$ be given by*

$$g(t) = \int_{-\infty}^{b(t)} f(x)\ dx; \quad h(t) = \int_{a(t)}^{\infty} f(x)\ dx,$$

where $a(t)$ and $b(t)$ are differentiable functions. Then $g(t)$ and $h(t)$ are differentiable, and

$$\begin{aligned} g'(t) &= f(b(t))\ b'(t); \\ h'(t) &= -f(a(t))\ a'(t). \end{aligned}$$

Proof. Since the integral of $f(x)$ over the entire real axis is finite, we can write the functions $g(t)$ and $h(t)$ as

$$\begin{aligned} g(t) &= \int_{-\infty}^{0} f(x)\ dx + \int_{0}^{b(t)} f(x)\ dx; \\ h(t) &= \int_{a(t)}^{0} f(x)\ dx + \int_{0}^{\infty} f(x)\ dx. \end{aligned}$$

Note that $\int_{-\infty}^{0} f(x)dx$ and $\int_{0}^{\infty} f(x)dx$ are constant numbers whose derivative with respect to t is 0. Then, using Lemma 1.2, we conclude that

$$\begin{aligned} g'(t) &= \frac{d}{dt}\left(\int_{0}^{b(t)} f(x)\ dx\right) = f(b(t))\ b'(t); \\ h'(t) &= \frac{d}{dt}\left(\int_{a(t)}^{0} f(x)\ dx\right) = -f(a(t))\ a'(t). \end{aligned}$$

□

2.4 Numerical methods for computing definite integrals: Midpoint rule, Trapezoidal rule, and Simpson's rule

Computing the value of a definite integral using the Fundamental Theorem of Calculus is only possible if a closed formula for the antiderivative of the function to integrate can be obtained. This is not always possible, e.g., for $\int e^{-x^2}\ dx$. In these cases, approximate values of the definite integral are computed using numerical integration methods. We present three of the most common such methods.

Let $f : [a, b] \to \mathbb{R}$ be an integrable function. To compute an approximate value of the integral

$$I = \int_a^b f(x)\ dx,$$

we partition the interval $[a, b]$ into n intervals of equal size $h = \frac{b-a}{n}$ by using the nodes $a_i = a + ih$, for $i = 0 : n$, i.e.,

$$a = a_0 < a_1 < a_2 < \ldots < a_{n-1} < a_n = b.$$

Note that $a_i - a_{i-1} = h$, $i = 1 : n$. Let x_i be the midpoint of the interval $[a_{i-1}, a_i]$, i.e.,

$$x_i = \frac{a_{i-1} + a_i}{2}, \quad \forall\ i = 1 : n.$$

The integral I can be written as

$$I = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} f(x)\ dx. \tag{2.12}$$

On each interval $[a_{i-1}, a_i]$, $i = 1 : n$, the function $f(x)$ is approximated by a simpler function whose integral on $[a_{i-1}, a_i]$ can be computed exactly. The resulting values are summed up to obtain an approximate value of I. Depending on whether constant functions, linear functions, or quadratic functions are used to approximate $f(x)$, the resulting numerical integration methods are called the Midpoint rule, the Trapezoidal rule, and the Simpson's rule, respectively.

Midpoint Rule: Approximate $f(x)$ on the interval $[a_{i-1}, a_i]$ by the constant function $c_i(x)$ equal to the value of the function f at the midpoint x_i of the interval $[a_{i-1}, a_i]$, i.e.,

$$c_i(x) = f(x_i), \quad \forall\ x \in [a_{i-1}, a_i]. \tag{2.13}$$

Then,

$$\int_{a_{i-1}}^{a_i} f(x)\ dx \approx \int_{a_{i-1}}^{a_i} c_i(x)\ dx = (a_i - a_{i-1}) f(x_i) = h f(x_i). \tag{2.14}$$

From (2.12) and (2.14), we obtain that the Midpoint Rule approximation I_n^M of I corresponding to n partition intervals is

$$\begin{aligned} I_n^M &= \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} c_i(x)\ dx \\ &= h \sum_{i=1}^{n} f(x_i). \end{aligned} \tag{2.15}$$

Trapezoidal Rule: Approximate $f(x)$ on the interval $[a_{i-1}, a_i]$ by the linear function $l_i(x)$ equal to $f(x)$ at the end points a_{i-1} and a_i, i.e.,

$$l_i(a_{i-1}) = f(a_{i-1}) \quad \text{and} \quad l_i(a_i) = f(a_i).$$

By linear interpolation, it is easy to see that

$$l_i(x) \;=\; \frac{x - a_{i-1}}{a_i - a_{i-1}}\, f(a_i) \;+\; \frac{a_i - x}{a_i - a_{i-1}}\, f(a_{i-1}), \quad \forall\, x \in [a_{i-1}, a_i]. \tag{2.16}$$

Then,

$$\int_{a_{i-1}}^{a_i} f(x)\, dx \;\approx\; \int_{a_{i-1}}^{a_i} l_i(x)\, dx \;=\; \frac{h}{2}\,(f(a_{i-1}) + f(a_i)). \tag{2.17}$$

From (2.12) and (2.17), we obtain that the Trapezoidal Rule approximation I_n^T of I corresponding to n partition intervals is

$$\begin{aligned} I_n^T &= \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} l_i(x)\, dx \\ &= \frac{h}{2}\left(f(a_0) + 2\sum_{i=1}^{n-1} f(a_i) + f(a_n) \right). \end{aligned} \tag{2.18}$$

Simpson's Rule: Approximate $f(x)$ on the interval $[a_{i-1}, a_i]$ by the quadratic function $q_i(x)$ equal to $f(x)$ at a_{i-1}, a_i, and at the midpoint $x_i = \frac{a_{i-1}+a_i}{2}$, i.e.,

$$q_i(a_{i-1}) = f(a_{i-1}); \quad q_i(x_i) = f(x_i) \quad \text{and} \quad q_i(a_i) = f(a_i).$$

By quadratic interpolation, we find that

$$\begin{aligned} q_i(x) \;=\; & \frac{(x - a_{i-1})(x - x_i)}{(a_i - a_{i-1})(a_i - x_i)} f(a_i) \;+\; \frac{(a_i - x)(x - a_{i-1})}{(a_i - x_i)(x_i - a_{i-1})} f(x_i) \\ & + \frac{(a_i - x)(x_i - x)}{(a_i - a_{i-1})(x_i - a_{i-1})} f(a_{i-1}), \quad \forall\, x \in [a_{i-1}, a_i]. \end{aligned} \tag{2.19}$$

Then,

$$\int_{a_{i-1}}^{a_i} f(x)\, dx \;\approx\; \int_{a_{i-1}}^{a_i} q_i(x)\, dx \;=\; \frac{h}{6}\,(f(a_{i-1}) + 4f(x_i) + f(a_i)). \tag{2.20}$$

From (2.12) and (2.20), we obtain that the Simpson's Rule approximation I_n^S of I corresponding to n partition intervals is

$$\begin{aligned} I_n^S &= \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} q_i(x)\, dx \\ &= \frac{h}{6}\left(f(a_0) + 2\sum_{i=1}^{n-1} f(a_i) + f(a_n) \;+\; 4\sum_{i=1}^{n} f(x_i) \right). \end{aligned} \tag{2.21}$$

Example: Compute the Midpoint, Trapezoidal, and Simpson's rules approximations of the definite integral

$$I \;=\; \int_1^3 \frac{1}{(x+1)^2}\, dx$$

for $n = 8$ partition intervals, and denote them by I_8^M, I_8^T, and I_8^S, respectively. Compute the exact value of I and find the approximation errors $|I - I_8^M|$, $|I - I_8^T|$, and $|I - I_8^S|$.

Answer: Let $f(x) = \frac{1}{(x+1)^2}$. For $n = 8$ partition intervals,

$$h = \frac{1}{4}; \quad a_i = 1 + \frac{i}{4},\ i = 0:8; \quad x_i = 1 + \frac{1}{4}\left(i - \frac{1}{2}\right),\ i = 1:8.$$

From (2.15), (2.18), and (2.21), respectively, we find that

$$\begin{aligned} I_8^M &= h\sum_{i=1}^{8} f(x_i) \;=\; \frac{1}{4}\sum_{i=1}^{8} \frac{1}{(1 + (i - 1/2)/4)^2} \;=\; 0.24943374; \\ I_8^T &= h\left(\frac{f(a_0)}{2} + \sum_{i=1}^{7} f(a_i) + \frac{f(a_8)}{2} \right) \;=\; \frac{1}{4}\left(\frac{1}{8} + \sum_{i=1}^{7} \frac{1}{(2 + i/4)^2} + \frac{1}{32} \right) \\ &= 0.25113543; \\ I_n^S &= \frac{h}{6}\left(f(a_0) + 2\sum_{i=1}^{7} f(a_i) + f(a_8) \;+\; 4\sum_{i=1}^{8} f(x_i) \right) \\ &= \frac{1}{24}\left(\frac{1}{4} + 2\sum_{i=1}^{7} \frac{1}{(2 + i/4)^2} + \frac{1}{16} \;+\; 4\sum_{i=1}^{8} \frac{1}{(1 + (i - 1/2)/4)^2} \right) \\ &= 0.25000097. \end{aligned}$$

It is easy to see that

$$I \;=\; \int_1^3 \frac{1}{(x+1)^2}\, dx \;=\; \left(-\frac{1}{x+1} \right)\Bigg|_1^3 \;=\; \frac{1}{4}.$$

The corresponding approximation errors are

$$\begin{aligned} |I - I_8^M| &= 0.00056625; \\ |I - I_8^T| &= 0.00113543; \\ |I - I_8^S| &= 0.00000097. \quad \square \end{aligned}$$

2.5 Convergence of the Midpoint, Trapezoidal, and Simpson's rules

In section 2.4, we derived the formulas (2.15), (2.18) and (2.21) for computing approximate values I_n^M, I_n^T and I_n^S of the integral

$$I = \int_a^b f(x)\,dx$$

corresponding to the Midpoint, Trapezoidal and Simpson's rules, respectively. However, it is not a priori clear whether these approximations are meaningful, i.e., whether I_n^M, I_n^T and I_n^S converge to I as n goes to infinity. In this section, we discuss the convergence of these methods.

Definition 2.1. *Denote by I_n the approximation of I obtained using a numerical integration method with n partition intervals. The method is convergent if and only if the approximations I_n converge to I as the number of intervals n goes to infinity (and therefore as $h = \frac{b-a}{n}$ goes to 0), i.e.,*

$$\lim_{n\to\infty} |I - I_n| = 0.$$

The order of convergence of the numerical integration method is $k > 0$ if and only if

$$|I - I_n| = O(h^k) = O\left(\frac{1}{n^k}\right).$$

Theorem 2.2. *Let $I = \int_a^b f(x)\,dx$, and let I_n^M, I_n^T, or I_n^S be the approximations of I given by the Midpoint, Trapezoidal, and Simpson's rules corresponding to n partition intervals of size $h = \frac{b-a}{n}$.*

(i) If $f''(x)$ exists and is continuous on $[a,b]$, then the approximation errors of the Midpoint and Trapezoidal rules can be bounded from above as follows:

$$|I - I_n^M| \leq \frac{h^2}{24}\,(b-a) \max_{a\leq x\leq b} |f''(x)|; \tag{2.22}$$

$$|I - I_n^T| \leq \frac{h^2}{12}\,(b-a) \max_{a\leq x\leq b} |f''(x)|. \tag{2.23}$$

Thus, the Midpoint and Trapezoidal rules are quadratically convergent, i.e.,

$$|I - I_n^M| = O\left(\frac{1}{n^2}\right), \quad \text{as } n \to \infty; \tag{2.24}$$

$$|I - I_n^T| = O\left(\frac{1}{n^2}\right), \quad \text{as } n \to \infty. \tag{2.25}$$

(ii) If $f^{(4)}(x)$ exists and is continuous on $[a,b]$, then

$$|I - I_n^S| \leq \frac{h^4}{2880}\,(b-a) \max_{a\leq x\leq b} |f^{(4)}(x)|, \tag{2.26}$$

and Simpson's rule is fourth order convergent, i.e.,

$$|I - I_n^S| = O\left(\frac{1}{n^4}\right), \quad \text{as } n \to \infty. \tag{2.27}$$

Summarizing the results of Theorem 2.2, if $f''(x)$ is continuous, then the Midpoint and Trapezoidal rules are second order convergent. Simpson's rule requires more smoothness of the function $f(x)$ for convergence, i.e., $f^{(4)}(x)$ must be continuous, but is then faster convergent, i.e., fourth order convergent, than the Midpoint and Trapezoidal rules.

Without giving a formal proof, we provide the intuition behind the results of Theorem 2.2.

The upper bounds (2.22), (2.23) and (2.26) can be established using the following approximation error results[1]: For any $i = 1 : n$, there exist points $\xi_{i,T}$, $\xi_{i,M}$ and $\xi_{i,S}$ in the interval $(a_{i-1},\ a_i)$ such that

$$\int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{a_{i-1}}^{a_i} c_i(x)\,dx = \frac{h^3}{24} f''(\xi_{i,T}); \tag{2.28}$$

$$\int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{a_{i-1}}^{a_i} l_i(x)\,dx = -\frac{h^3}{12} f''(\xi_{i,M}); \tag{2.29}$$

$$\int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{a_{i-1}}^{a_i} q_i(x)\,dx = -\frac{h^5}{2880} f^{(4)}(\xi_{i,S}), \tag{2.30}$$

where the functions $c_i(x)$, $l_i(x)$, and $q_i(x)$ are given by (2.13), (2.16), and (2.19), respectively.

For, e.g., the Midpoint rule, we find from (2.12), (2.15), and (2.28) that

$$\begin{aligned}
|I - I_n^M| &= \left| \sum_{i=1}^n \int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{a_{i-1}}^{a_i} c_i(x)\,dx \right| \\
&\leq \sum_{i=1}^n \left| \int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{a_{i-1}}^{a_i} c_i(x)\,dx \right| \\
&= \frac{h^3}{24} \sum_{i=1}^n |f''(\xi_{i,T})| = \frac{h^2}{24}\,(b-a)\,\frac{\sum_{i=1}^n |f''(\xi_{i,T})|}{n}
\end{aligned}$$

[1]The approximation results (2.28–2.30) can be derived either by using general properties of interpolating polynomials, or by using Taylor approximations, provided that the function $f(x)$ has the smoothness required in Theorem 2.2.

$$\leq \frac{h^2}{24}(b-a)\max_{a\leq x\leq b}|f''(x)|,$$

since $|f''(\xi_{i,T})| \leq \max_{a\leq x\leq b}|f''(x)|$, for all $i = 1 : n$.

Formula (2.22) is therefore proven. The proofs of (2.23) and (2.26), follow from similarly from (2.29) and (2.30).

The convergence results (2.24), (2.25) and (2.27) follow easily from (2.22), (2.23) and (2.26). Recall that any continuous function on a closed interval has a finite maximum, which is achieved at (at least) one point of the interval. If $f''(x)$ and $f^{(4)}(x)$, respectively, are continuous on $[a,b]$, let

$$M_2 = \max_{a\leq x\leq b}|f''(x)| < \infty \quad \text{and} \quad M_4 = \max_{a\leq x\leq b}|f^{(4)}(x)| < \infty.$$

Since $h = \frac{b-a}{n}$, it follows that (2.22), (2.23), and (2.26) can be written as

$$\begin{aligned}
|I - I_n^M| &\leq \frac{1}{n^2}\frac{(b-a)^3M_2}{24};\\
|I - I_n^T| &\leq \frac{1}{n^2}\frac{(b-a)^3M_2}{12};\\
|I - I_n^S| &\leq \frac{1}{n^4}\frac{(b-a)^5M_4}{2880},
\end{aligned}$$

respectively, and therefore (2.24), (2.25) and (2.27) are established.

2.5.1 Implementation of numerical integration methods

Computing approximate values of the definite integral of a given function $f(x)$ on the interval $[a,b]$, using the Midpoint, Trapezoidal or Simpson's rules requires the implementation of formulas (2.15), (2.18), and (2.21), i.e.,

$$I_n^M = h\sum_{i=1}^{n} f(x_i); \tag{2.31}$$

$$I_n^T = h\left(\frac{f(a_0)}{2} + \frac{f(a_n)}{2}\right) + h\sum_{i=1}^{n-1} f(a_i); \tag{2.32}$$

$$I_n^S = h\left(\frac{f(a_0)}{6} + \frac{f(a_n)}{6}\right) + \frac{h}{3}\sum_{i=1}^{n-1} f(a_i) + \frac{2h}{3}\sum_{i=1}^{n} f(x_i). \tag{2.33}$$

Here, $h = \frac{b-a}{n}$, $a_i = a + ih$, $i = 0 : n$, and $x_i = a + \left(i - \frac{1}{2}\right)h$, $i = 1 : n$.

A routine f_int(x) evaluating the function to be integrated at the point x is required. The end points a and b of the integration interval and the number of intervals n must also be specified.

The pseudocodes for the Midpoint, Trapezoidal, and Simpson's rules are given in Tables 2.1, 2.2, and 2.3, respectively.

Table 2.1: Pseudocode for Midpoint Rule

```
Input:
a = left endpoint of the integration interval
b = right endpoint of the integration interval
n = number of partition intervals
f_int(x) = routine evaluating f(x)

Output:
I_midpoint = Midpoint Rule approximation of ∫_a^b f(x)

h = (b − a)/n;  I_midpoint = 0
for i = 1 : n
  I_midpoint = I_midpoint + f_int(a + (i − 1/2)h)
end
I_midpoint = h · I_midpoint
```

Table 2.2: Pseudocode for Trapezoidal Rule

```
Input:
a = left endpoint of the integration interval
b = right endpoint of the integration interval
n = number of partition intervals
f_int(x) = routine evaluating f(x)

Output:
I_trap = Trapezoidal Rule approximation of ∫_a^b f(x)

h = (b − a)/n
I_trap = f_int(a)/2 + f_int(b)/2
for i = 1 : (n − 1)
  I_trap = I_trap + f_int(a + ih)
end
I_trap = h · I_trap
```

In practice, we want to find an approximate value that is within a prescribed tolerance *tol* of the integral I of a given function $f(x)$ over the interval

Table 2.3: Pseudocode for Simpson's Rule

```
Input:
a = left endpoint of the integration interval
b = right endpoint of the integration interval
n = number of partition intervals
f_int(x) = routine evaluating f(x)

Output:
I_simpson = Simpson's Rule approximation of ∫_a^b f(x)

h = (b − a)/n
I_simpson = f_int(a)/6 + f_int(b)/6
for i = 1 : (n − 1)
  I_simpson = I_simpson + f_int(a + ih)/3
end
for i = 1 : n
  I_simpson = I_simpson + 2 f_int(a + (i − 1/2)h)/3
end
I_simpson = h · I_simpson
```

$[a, b]$. Simply using a numerical integration methods with n partition intervals cannot work effectively, since we do not know in advance how large should n be chosen to obtain an approximation of I with the desired accuracy.

What we do is choose an integration method and a small number of intervals, e.g., 4 or 8 intervals, and compute the numerical approximation of the integral. We then double the number of intervals and compute another approximation of I. If the absolute value of the difference between the new and old approximations is smaller than the required tolerance *tol*, we declare the last computed approximation of the integral to be the approximate value of I that we are looking for. Otherwise, the number of intervals is again doubled and the process is repeated until two consecutive numerical integration approximations are within the desired tolerance *tol* of each other. This condition is called the stopping criterion for the algorithm, and can be written formally as

$$|I_{new} - I_{old}| \; < \; tol, \tag{2.34}$$

where I_{old} and I_{new} are the last two approximations of I that were computed (and therefore I_{new} corresponds to twice as many intervals as I_{old}).

The pseudocode for this method is given in Table 2.4.

It is interesting to note that, if the Trapezoidal rule or Simpson's rule are used, about half of the nodes from the partition of the interval $[a, b]$ with $2n$ intervals are also nodes in the partition of $[a, b]$ with n intervals. Thus,

Table 2.4: Pseudocode for computing an approximate value of an integral with given tolerance

```
Input:
tol = prescribed tolerance
I_numerical(n) = result of the numerical integration rule with
    n intervals; any integration rule can be used

Output:
I_approx = approximation of ∫_a^b f(x) with tolerance tol

n = 4;  I_old = I_numerical(n)        // 4 intervals initial partition
n = 2n;  I_new = I_numerical(n)
while (abs(I_new − I_old) > tol)
  I_old = I_new
  n = 2n
  I_new = I_numerical(n)
end
I_approx = I_new
```

when computing the approximations I_{2n}^T or I_{2n}^S, we do not need to evaluate the function $f(x)$ at all nodes, provided that the values of $f(x)$ required for computing the approximations I_n^T or I_n^S were stored separately. Depending on how computationally expensive it is to evaluate $f(x)$ at a point x, this might result in computational savings.

For example, when using the trapezoidal method, we find from (2.18) that

$$I_n^T \;=\; \frac{h_n}{2}\left(f(a_{n,0}) + 2\sum_{i=1}^{n-1} f(a_{n,i}) + f(a_{n,n})\right),$$

where $h_n = \frac{b-a}{n}$ and $a_{n,i} = a + ih_n$, for $i = 0 : n$.

When the number of intervals is doubled from n to $2n$, we obtain that

$$I_{2n}^T \;=\; \frac{h_{2n}}{2}\left(f(a_{2n,0}) + 2\sum_{i=1}^{2n-1} f(a_{2n,i}) + f(a_{2n,2n})\right),$$

where $h_{2n} = \frac{b-a}{2n}$ and $a_{2n,i} = a + ih_{2n}$, for $i = 0 : 2n$. Note that

$$a_{2n,2i} = a + 2ih_{2n} = a + 2i\frac{b-a}{2n} = a + i\frac{b-a}{n} = a + ih_n = a_{n,i}, \quad \forall\ i = 0 : n.$$

In other words, if all the values of $f(a_{n,i})$, $i = 0 : n$, computed when evaluating I_n^T are saved, then only the n values of $f(a_{2n,2i+1})$, $i = 0 : (n-1)$, must be

computed when evaluating I_{2n}^T, since $f(a_{2n,2i}) = f(a_{n,i})$, for $i = 0 : n$. (Note that additional storage costs are nonetheless incurred.)

Similar savings occur for Simpson's rule. However, there are no such savings to be obtained for the Midpoint rule.

2.5.2 A concrete example

We present an example of how to compute an approximate value of a given integral to a prescribed tolerance. We want to find an approximate value for

$$I = \int_0^2 e^{-x^2}\,dx$$

which is within $0.5\ 10^{-7}$ of I. Note that an exact value of I cannot be computed, since $\int e^{-x^2}\,dx$ does not have a closed formula.

We implement the algorithm from Table 2.4 for each of the numerical integration methods to compare their convergence properties. We choose $tol = 0.5\ 10^{-7}$. For an initial partition of the interval $[0,\ 2]$ into $n = 4$ intervals, the following approximate values of I are found using the Midpoint, Trapezoidal, and Simpson's rules, respectively:

$$I_4^M = 0.88278895; \quad I_4^T = 0.88061863; \quad I_4^S = 0.88206551.$$

Then, we double the number of partition intervals and compute the numerical approximations corresponding to each method. We keep doubling the number of partition intervals until the stopping criterion (2.34) is satisfied. The results are recorded below:

No. Intervals	Midpoint Rule	Trapezoidal Rule	Simpson's Rule
4	0.88278895	0.88061863	0.88206551
8	0.88226870	0.88170379	0.88208040
16	0.88212887	0.88198624	0.88208133
32	0.88209330	0.88205756	0.88208139
64	0.88208437	0.88207543	
128	0.88208214	0.88207990	
256	0.88208158	0.88208102	
512	0.88208144	0.88208130	

We note that convergence is achieved for 512 intervals for the Midpoint and Trapezoidal rules, and for 32 intervals for Simpson's rule.

To better understand the convergence patterns of the quadratically convergent Midpoint and Trapezoidal rules, and of the fourth order convergent

Simpson's rule, we look at the approximation errors of each algorithm. Since an exact value of I cannot be computed, we assume that the approximate value obtained using Simpson's rule with 100,000 intervals to be the exact value of I, i.e.,

$$I = 0.88208139076242.$$

The approximation errors $|I - I_n^M|$, $|I - I_n^T|$, and $|I - I_n^S|$ for the Midpoint, Trapezoidal, and Simpson's rules, respectively, are presented below:

No. Intervals	Midpoint Rule	Trapezoidal Rule	Simpson's Rule
n	$\lvert I - I_n^M\rvert$	$\lvert I - I_n^T\rvert$	$\lvert I - I_n^S\rvert$
4	0.00070756	0.00146276	0.00001588
8	0.00018731	0.00037760	0.00000099
16	0.00004748	0.00009515	0.00000006
32	0.00001191	0.00002383	$3.88/10^9$
64	0.00000298	0.00000596	
128	0.00000075	0.00000149	
256	0.00000019	0.00000037	
512	0.00000005	0.00000009	

Note that the approximation errors for the Midpoint rule are about half of the approximation errors for the Trapezoidal rule. While this is not always the case, it is nonetheless consistent with the theoretical upper bounds (2.22) and (2.23) from Theorem 2.2, i.e.,

$$|I - I_n^M| \leq \frac{h^2}{24}\,(b-a) \max_{a\leq x\leq b} |f''(x)|;$$

$$|I - I_n^T| \leq \frac{h^2}{12}\,(b-a) \max_{a\leq x\leq b} |f''(x)|.$$

As the number of intervals doubles, the approximation error decreases by a factor of 4 for the Midpoint and Trapezoidal rules, and by a factor of 16 for Simpson's rule. This is consistent with the results (2.24), (2.25) and (2.27) from Theorem 2.2, i.e.,

$$|I - I_n^M| = O\left(\frac{1}{n^2}\right); \quad |I - I_n^T| = O\left(\frac{1}{n^2}\right); \quad |I - I_n^S| = O\left(\frac{1}{n^4}\right).$$

As in the example above, in most cases, Simpson's rule converges faster than the Trapezoidal and Midpoint rules. Nonetheless, from a computational point of view, it is more expensive to compute the Simpson's rule approximation I_n^S, which requires evaluating the function $f(x)$ at $2n+1$ nodes, that to compute the Trapezoidal rule approximation I_n^T, which requires evaluating

the function $f(x)$ at only $n+1$ nodes. Depending on how computationally intensive it is to evaluate the function $f(x)$ at a given point x, it may happen that the trapezoidal rule is more accurate (i.e., produces a smaller approximation error) than Simpson's rule for comparable computing costs.

FINANCIAL APPLICATIONS

Interest Rate Curves. Zero rates and instantaneous interest rates. Forward rates. Continuously and discretely compounded interest.

Bond Pricing. Yield of a Bond. Bond Duration and Bond Convexity.

Zero coupon bonds.

Numerical implementation of bond mathematics.

2.6 Interest Rate Curves. Zero rates and instantaneous rates

The zero rate $r(0,t)$ between time 0 and time t is the rate of return of a cash deposit made at time 0 and maturing at time t. If specified for all values of t, then $r(0,t)$ is called the zero rate curve[2] and is a continuous function of t.

Interest can be compounded at discrete time intervals, e.g., annually, semiannually, monthly, etc., or can be compounded continuously. Unless otherwise specified, we assume that interest is compounded continuously. For continuously compounded interest, the value at time t of $B(0)$ currency units (e.g., U.S. dollars) is

$$B(t) \;=\; \exp(t\, r(0,t))\; B(0), \tag{2.35}$$

where $\exp(x) = e^x$. The value at time 0 of $B(t)$ currency units at time t is

$$B(0) \;=\; \exp(-t\, r(0,t))\; B(t). \tag{2.36}$$

The instantaneous rate $r(t)$ at time t is the rate of return of deposits made at time t and maturing at time $t+dt$, where dt is infinitesimally small, i.e.,

$$r(t) \;=\; \lim_{dt\to 0} \frac{1}{dt} \cdot \frac{B(t+dt)-B(t)}{B(t)} \;=\; \frac{B'(t)}{B(t)}.$$

[2]We note, and further explain this in section 2.7.1, that $r(0,t)$ is the yield of a zero-coupon bond with maturity t. The zero rate curve is also called the yield curve.

We conclude that $B(t)$ satisfies the ordinary differential equation (ODE)

$$\frac{B'(\tau)}{B(\tau)} \;=\; r(\tau), \quad \forall\, \tau > 0, \tag{2.37}$$

with the initial condition that $B(\tau)$ at time $\tau = 0$ must be equal to $B(0)$. By integrating (2.37) between 0 and $t > 0$, it follows that

$$\int_0^t r(\tau)\; d\tau \;=\; \int_0^t \frac{B'(\tau)}{B(\tau)}\; d\tau \;=\; \ln(B(\tau))\Big|_{\tau=0}^{\tau=t} \;=\; \ln\left(\frac{B(t)}{B(0)}\right).$$

Therefore,

$$B(t) \;=\; B(0)\, \exp\left(\int_0^t r(\tau)\; d\tau\right), \quad \forall\, t > 0. \tag{2.38}$$

Formula (2.38) gives the future value at time t of a cash deposit made at time $0 < t$. It can also be used to find the present value at time 0 of a cash deposit $B(t)$ made at time $t > 0$, i.e.,

$$B(0) \;=\; B(t)\, \exp\left(-\int_0^t r(\tau)\; d\tau\right), \quad \forall\, t > 0; \tag{2.39}$$

The term $\exp\left(-\int_0^t r(\tau)\; d\tau\right)$ from (2.39) is called the discount factor.

From (2.35) and (2.38), it follows that

$$r(0,t) \;=\; \frac{1}{t}\int_0^t r(\tau)\; d\tau. \tag{2.40}$$

In other words, the zero rate $r(0,t)$ is the average of the instantaneous rate $r(t)$ over the time interval $[0,t]$.

If $r(t)$ is continuous, then it is uniquely determined if the zero rate curve $r(0,t)$ is known. From (2.40), we obtain that

$$\int_0^t r(\tau)\; d\tau \;=\; t\, r(0,t). \tag{2.41}$$

By differentiating (2.41) with respect to t, see, e.g., Lemma 1.2, we find that

$$r(t) \;=\; r(0,t) \;+\; t\,\frac{d}{dt}(r(0,t)). \tag{2.42}$$

Formulas (2.38) and (2.39) can also be written for times $t_1 < t_2$, instead of times 0 and t, as follows:

$$B(t_2) \;=\; B(t_1)\, \exp\left(\int_{t_1}^{t_2} r(\tau)\; d\tau\right), \quad \forall\, 0 < t_1 < t_2; \tag{2.43}$$

$$B(t_1) \;=\; B(t_2)\, \exp\left(-\int_{t_1}^{t_2} r(\tau)\; d\tau\right), \quad \forall\, 0 < t_1 < t_2. \tag{2.44}$$

2.6.1 Constant interest rates

A common assumption made when pricing derivative securities (with relatively short maturities) is that the risk–free rates are constant over the life of the derivative security. (Such an assumption is used, e.g., in the Black–Scholes model; see section 3.5 for more details.)

In section 2.6, we introduced two different types of interest rates: the zero rates $r(0,t)$ and the instantaneous rates $r(t)$. However, when the assumption that interest rates are constant is made, it is never mentioned whether the zero rates or the instantaneous rates are considered to be constant. The reason is that if either one of these rates is constant and equal to r, then the other rate is also constant and equal to r, i.e.,

$$r(0,t) = r, \ \forall\, 0 \le t \le T \quad \Longleftrightarrow \quad r(t) = r, \ \forall\, 0 \le t \le T.$$

To see this, assume that $r(t) = r$, $\forall\, 0 \le t \le T$. From (2.40), we find that

$$r(0,t) \;=\; \frac{1}{t}\int_0^t r(\tau)\, d\tau \;=\; \frac{1}{t}\int_0^t r\, d\tau \;=\; r, \quad \forall\, 0 < t \le T.$$

Since $r(0,t)$ is continuous, we obtain that $r(0,0) = r$.

If $r(0,t) = r$, $\forall\, 0 \le t \le T$, recall from (2.42) that

$$r(t) \;=\; r(0,t) \;+\; t\,\frac{d}{dt}(r(0,t)) \;=\; r, \quad \forall\, 0 < t < T.$$

Then $r(0) = r$ and $r(T) = r$ as well, since $r(t)$ is continuous.

If interest rates are constant and equal to r, the future value and present value formulas (2.38), (2.39), (2.43), and (2.44) become

$$\begin{aligned}
B(t) &= e^{rt}\, B(0), \quad \forall\, t > 0; && (2.45)\\
B(t_2) &= e^{r(t_2-t_1)}\, B(t_1), \quad \forall\, 0 < t_1 < t_2; && (2.46)\\
B(0) &= e^{-rt}\, B(t), \quad \forall\, t > 0; && (2.47)\\
B(t_1) &= e^{-r(t_2-t_1)}\, B(t_2), \quad \forall\, 0 < t_1 < t_2. && (2.48)
\end{aligned}$$

2.6.2 Forward Rates

The forward rate of return $r(0;t_1,t_2)$ between times t_1 and t_2 is the constant rate of return, *as seen at time 0*, of a deposit that will be made at time $t_1 > 0$ in the future and will mature at time $t_2 > t_1$.

An arbitrage–free value for the forward rate $r(0;t_1,t_2)$ in terms of the zero rate curve $r(0,t)$ can be found using the Law of One Price as follows. Consider two different strategies for investing $B(0)$ currency units at time 0:

First Strategy: At time 0, deposit $B(0)$ currency units until time t_1, with interest rate $r(0,t_1)$. Then, at t_1, deposit the proceeds until time t_2, at the forward rate $r(0;t_1,t_2)$, which was locked in at time 0. The value of the deposit at time t_1 is $V_1(t_1) = B(0)\exp(t_1 r(0,t_1))$. At time t_2, the value is

$$\begin{aligned}
V_1(t_2) &= V_1(t_1)\,\exp(\,(t_2-t_1)\, r(0;t_1,t_2)\,)\\
&= B(0)\,\exp(\,t_1\, r(0,t_1) \;+\; (t_2-t_1)\, r(0;t_1,t_2)\,). && (2.49)
\end{aligned}$$

Second Strategy: Deposit $B(0)$ at time 0 until time t_2, with interest rate $r(0,t_2)$. At time t_2, the value of the deposit is

$$V_2(t_2) \;=\; B(0)\,\exp(t_2\, r(0,t_2)). \qquad (2.50)$$

Both investment strategies are risk free and the cash amount invested at time 0 is the same, equal to $B(0)$, for both strategies. From the Law of One Price, see Theorem 1.10, it follows that $V_1(t_2) = V_2(t_2)$. From (2.49) and (2.50), we find that

$$t_1\, r(0,t_1) \;+\; (t_2-t_1)\, r(0;t_1,t_2) \;=\; t_2\, r(0,t_2).$$

By solving for $r(0;t_1,t_2)$, we conclude that

$$r(0;t_1,t_2) \;=\; \frac{t_2\, r(0,t_2) \;-\; t_1\, r(0,t_1)}{t_2 - t_1}. \qquad (2.51)$$

Note that, by definition, the forward rate $r(0;t_1,t_2)$ is the instantaneous rate $r(t_1)$ in the limiting case when t_2 goes to t_1, i.e., $\lim_{t_2 \searrow t_1} r(0;t_1,t_2) = r(t_1)$. Then, by taking the limit as t_2 goes to t_1 in (2.51) we should obtain formula (2.42) providing the connection between $r(t)$ and $r(0,t)$. This is indeed the case:

$$\begin{aligned}
r(t_1) &= \lim_{t_2 \searrow t_1} r(0;t_1,t_2) \;=\; \lim_{t_2 \searrow t_1} \frac{t_2\, r(0,t_2) \;-\; t_1\, r(0,t_1)}{t_2 - t_1}\\
&= \lim_{t_2 \searrow t_1}\left(r(0,t_2) \;+\; \frac{t_1\,(r(0,t_2) - r(0,t_1))}{t_2 - t_1}\right)\\
&= r(0,t_1) \;+\; t_1\,\frac{d}{dt_1}(r(0,t_1)),
\end{aligned}$$

which is the same as (2.42).

2.6.3 Discretely compounded interest

In most of the book we assume that interest is compounded continuously. However, this is not always (or even often) the case in the markets. In this section, we present here several types of discretely compounded interests.

Assume that interest is compounded n times every year, and let $r_n(0,t)$, for $t \geq 0$, denote the corresponding zero rate curve. Then,

$$B_n(t) \;=\; B(0)\left(1+\frac{r_n(0,t)}{n}\right)^{nt}, \tag{2.52}$$

where $B_n(t)$ is the amount that accumulates at time t from an amount $B(0)$ at time 0 by compounding interest n times a year between 0 and t.

The most common types of discretely compounded interest are:

- annually compounded, i.e., once a year:

$$B_1(t) \;=\; B(0)\;(1+r_1(0,t))^t;$$

- semiannually compounded, i.e., every six months:

$$B_2(t) \;=\; B(0)\left(1+\frac{r_2(0,t)}{2}\right)^{2t};$$

- quarterly compounded, i.e., every three months:

$$B_4(t) \;=\; B(0)\left(1+\frac{r_4(0,t)}{4}\right)^{4t};$$

- monthly compounded, i.e., every month:

$$B_{12}(t) \;=\; B(0)\left(1+\frac{r_{12}(0,t)}{12}\right)^{12t}.$$

Note that continuously compounded interest is the limiting case of discretely compounded interest, if interest is compounded very frequently (e.g., every minute or every second), i.e., $B_{cont}(t) = \lim_{n\to\infty} B_n(t)$. Here, $B_n(t)$ and $B_{cont}(t)$ are given by (2.52) and (2.35), respectively, i.e., $B_{cont}(t) = e^{tr(0,t)}B(0)$.

To see this, assume that, as $n \to \infty$, the interest rate curves $r_n(0,t)$ are uniformly bounded and converge to a function, denoted by $r(0,t)$, i.e.,

$$r(0,t) \;=\; \lim_{n\to\infty} r_n(0,t).$$

Then, from (2.52), we obtain that

$$\begin{aligned}
\lim_{n\to\infty} B_n(t) &= \lim_{n\to\infty} B(0)\left(1\;+\;\frac{r_n(0,t)}{n}\right)^{nt}\\
&= \lim_{n\to\infty} B(0)\left(\left(1+\frac{r_n(0,t)}{n}\right)^{\frac{n}{r_n(0,t)}}\right)^{t\,r_n(0,t)}\\
&= B(0)\;e^{t\;r(0,t)} \;=\; B_{cont}(t).
\end{aligned}$$

Here, we used the fact that $\frac{n}{r_n(0,t)} \to \infty$ as $n \to \infty$, since $r_n(0,t)$ is uniformly bounded for any fixed t, and that $e = \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$; cf. (1.32).

2.7 Bond Pricing. Yield of a Bond. Bond Duration and Bond Convexity

A bond is a financial instrument used to issue debt, i.e., to borrow money. The issuer of the bond receives cash when the bond is issued, and must pay the face value of the bond (also called the principal of the bond) at a certain time in the future called the maturity (or expiry) of the bond. Other payments by the issuer of the bond to the buyer of the bond, called coupon payments, may also be made at predetermined times in the future. (A coupon payment is also made at maturity in this case.) The price of the bond is equal to the sum of all the future cash flows discounted to the present by using the risk-free zero rate.

Let B be the value of a bond with future cash flows c_i to be paid to the holder of the bond at times t_i, $i = 1:n$. Let $r(0,t_i)$ be the continuously compounded zero rates corresponding to t_i, $i = 1:n$. Then,

$$B \;=\; \sum_{i=1}^{n} c_i e^{-r(0,t_i)t_i}. \tag{2.53}$$

Note that $e^{-r(0,t_i)t_i}$ is the discount factor corresponding to time t_i, $i = 1:n$.

If the instantaneous interest rate curve $r(t)$ is known, a formula similar to (2.53) can be given, by substituting the discount factor $\exp\left(-\int_0^{t_i} r(\tau)\,d\tau\right)$ for $e^{-r(0,t_i)t_i}$; see (2.40). Then,

$$B \;=\; \sum_{i=1}^{n} c_i \exp\left(-\int_0^{t_i} r(\tau)\;d\tau\right). \tag{2.54}$$

Example: A semiannual coupon bond, with face value F, coupon rate C, and maturity T (measured in years), pays the holder of the bond a coupon payment equal to $\frac{C}{2}\,F$ every six months, except at maturity. The final payment at the maturity T is equal to one coupon payment plus the face value of the bond, i.e., $F\;+\;\frac{C}{2}\,F$. Let $t_1,\,t_2,\,\ldots,\,t_{n-1},\,t_n = T$ be the coupon dates (the dates when the payments are due)[3]. Let $r(0,t_i)$ be the zero rates

[3]Note that, for a semiannual coupon bond, $t_{i+1} = t_i + 0.5$, $\forall\; i = 1:(n-1)$, since

corresponding to time t_i, $i = 1 : n$. According to formula (2.53), the price B of the bond is computed as follows:

$$B = \sum_{i=1}^{n-1} \frac{C}{2} F\, e^{-r(0,t_i)t_i} + \left(F + \frac{C}{2} F\right) e^{-r(0,T)T}. \quad \square \qquad (2.55)$$

If all the cash flows c_i are proportional to the face value of the bond, then the price of the bond is proportional to its face value. Therefore, it is usually assumed that the face value of a bond is equal to 100, i.e., $F = 100$. We will follow this convention as well, unless otherwise specified.

Definition 2.2. *The yield of a bond is the internal rate of return of the bond, i.e., the constant rate at which the sum of the discounted future cash flows of the bond is equal to the price of the bond. If B is the price of a bond with cash flows c_i at time t_i, $i = 1 : n$, and if y is the yield of the bond, then,*

$$B = \sum_{i=1}^{n} c_i e^{-yt_i}. \qquad (2.56)$$

As expressed in (2.56), the price of the bond B can be regarded as a function of the yield. Therefore, whenever needed, we think of B as being a function of the yield, i.e., $B = B(y)$.

It is easy to see that the price of the bond goes down if the yield goes up, and it goes up if the yield goes down.

To compute the yield of a bond with a known price B, we must solve (2.56) for y. This can be written as a nonlinear equation in y, i.e.,

$$f(y) = 0, \quad \text{where} \quad f(y) = \sum_{i=1}^{n} c_i e^{-yt_i} - B,$$

which is then solved numerically by, e.g., Newton's Method; see section 8.5 for more details on computing bond yields.

Definition 2.3. *Par yield is the coupon rate that makes the value of the bond equal to its face value*[4].

In other words, par yield is the value C of the coupon rate such that $B = F$. If interest rates are positive, then the par yield is uniquely determined.

payments are made every six months. For example, for a semiannual coupon bond with 15 months to maturity, there are three coupon dates, in 3, 9, and 15 months, corresponding to $t_1 = \frac{3}{12} = \frac{1}{4}$, $t_2 = \frac{9}{12} = \frac{3}{4}$, and $t_3 = \frac{15}{12} = \frac{5}{4}$, respectively.

[4]For semiannually compounded interest, it can be shown that the yield of a bond with coupon rate equal to the par yield is exactly equal to the par yield.

For the semiannual coupon bond considered previously, we substitute $B = F$ in (2.55). Then, the par yield of the bond can be obtained by solving the following linear equation for C:

$$1 = \sum_{i=1}^{n-1} \frac{C}{2} e^{-r(0,t_i)t_i} + \left(1 + \frac{C}{2}\right) e^{-r(0,T)T}.$$

Duration and convexity are two of the most important parameters to estimate when investing in a bond, other than its yield. Duration provides the sensitivity of the bond price with respect to small changes in the yield, while convexity distinguishes between two bond portfolios with the same duration. (The portfolio with higher convexity is more desirable.)

The duration[5] of a bond is the weighted time average of the future cash flows of the bond discounted with respect to the yield of the bond, and normalized by dividing by the price of the bond.

Definition 2.4. *The duration D of a bond with price B and yield y, with cash flows c_i at time t_i, $i = 1 : n$, is*

$$D = \frac{\sum_{i=1}^{n} t_i c_i e^{-yt_i}}{B}. \qquad (2.57)$$

From (2.56) and (2.57), it is easy to see that

$$\frac{\partial B}{\partial y} = -\sum_{i=1}^{n} t_i c_i e^{-yt_i} = -B\,D,$$

and therefore

$$D = -\frac{1}{B}\frac{\partial B}{\partial y}. \qquad (2.58)$$

The duration of a bond gives the relative change in the price of a bond for *small* changes Δy in the yield of the bond (also known as parallel shifts of the yield curve). Let ΔB be the corresponding change in the price of the bond, i.e., $\Delta B = B(y + \Delta y) - B(y)$. The discretized version[6] of (2.58) is

$$D \approx -\frac{1}{B}\frac{B(y+\Delta y) - B(y)}{\Delta y} = -\frac{\Delta B}{B \cdot \Delta y},$$

[5]The duration from (2.57) corresponds to continuously compounded yield and is called Macaulay duration. If the yield is compounded discretely, e.g., m times a year, then modified duration is defined as the Macaulay duration divided by $1 + \frac{y}{m}$, where y is the yield of the bond. Formula (2.58) also holds for modified duration.

[6]The approximation $\frac{\partial B}{\partial y} \approx \frac{B(y+\Delta y)-B(y)}{\Delta y}$ can be regarded as a first order finite difference approximation of the first derivative of $B(y)$ with respect to y; see, e.g., (6.3).

which is equivalent to

$$\frac{\Delta B}{B} \approx -\Delta y\ D. \tag{2.59}$$

In other words, the percentage change in the price of the bond can be approximated by the duration of the bond multiplied by the parallel shift in the yield curve, with opposite sign.

For very small parallel shifts in the yield curve, the approximation formula (2.59) is accurate. For larger parallel shifts, convexity is used to better capture the effect of the changes in the yield curve on the price of the bond.

Definition 2.5. *The convexity C of a bond with price B and yield y is*

$$C = \frac{1}{B}\frac{\partial^2 B}{\partial y^2}. \tag{2.60}$$

Using (2.56), it is easy to see that

$$C = \frac{\sum_{i=1}^n t_i^2 c_i e^{-yt_i}}{B}. \tag{2.61}$$

The following approximation of the percentage change in the price of the bond for a given a change in the yield of the bond is more accurate than (2.59) and will be proved in section 5.6 using Taylor expansions:

$$\frac{\Delta B}{B} \approx -D\Delta y + \frac{1}{2}C(\Delta y)^2.$$

2.7.1 Zero Coupon Bonds

A zero coupon bond is a bond that pays back the face value of the bond at maturity and has no other payments, i.e., has coupon rate equal to 0. If F is the face value of a zero coupon bond with maturity T, the bond pricing formula (2.53) becomes

$$B = F\ e^{-r(0,T)T}, \tag{2.62}$$

where B is the price of the bond at time 0 and $r(0,T)$ is the zero rate corresponding to time T.

If the instantaneous interest rate curve $r(t)$ is given, the bond pricing formula (2.54) becomes

$$B = F \exp\left(-\int_0^T r(\tau)\ d\tau\right). \tag{2.63}$$

Let y be the yield of the bond. From (2.56), we find that

$$B = Fe^{-yT}. \tag{2.64}$$

From (2.62) and (2.64), we conclude that $y = r(0,T)$. In other words, the yield of a zero coupon bond is the same as the zero rate corresponding to the maturity of the bond. This explains why the zero rate curve $r(0,t)$ is also called the yield curve.

As expected, the duration of a zero coupon bond is equal to the maturity of the bond. From (2.58) and (2.64), we obtain that

$$D = -\frac{1}{B}\frac{\partial B}{\partial y} = -\frac{1}{Fe^{-yT}}\left(-T\ Fe^{-yT}\right) = T.$$

The convexity of a zero coupon bond can be computed from (2.60) and (2.64):

$$C = \frac{1}{B}\frac{\partial^2 B}{\partial y^2} = \frac{1}{Fe^{-yT}}\left(T^2\ Fe^{-yT}\right) = T^2.$$

2.8 Numerical implementation of bond mathematics

When specifying a bond, the maturity T of the bond, as well as the cash flows c_i and the cash flows dates t_i, $i = 1 : n$, are given. The price of the bond can be obtained from formula (2.53), i.e.,

$$B = \sum_{i=1}^n c_i e^{-r(0,t_i)t_i},$$

provided that the zero rate curve $r(0,t)$ is known for any $t > 0$, or at least for the cash flow times, i.e., for $t = t_i$, $i = 1 : n$.

If a routine r_zero(t) for computing the zero rate curve is given, the pseudocode from Table 2.5 can be used to compute the price of the bond.

If the instantaneous interest rate curve $r(t)$ is known, the price of the bond is given by formula (2.54), i.e.,

$$B = \sum_{i=1}^n c_i \exp\left(-\int_0^{t_i} r(\tau)\ d\tau\right). \tag{2.65}$$

If a closed formula for $\int r(\tau)\ d\tau$ cannot be found, evaluating the discount factors $disc(i) = \exp\left(-\int_0^{t_i} r(\tau)\ d\tau\right)$, $i = 1 : n$, requires estimating

$$I_i = \int_0^{t_i} r(\tau)\ d\tau, \quad \forall\ i = 1 : n,$$

Table 2.5: Pseudocode for computing the bond price given the zero rate curve

```
Input:
n = number of cash flows
t_cash_flow = vector of cash flow dates (of size n)
v_cash_flow = vector of cash flows (of size n)
r_zero(t) = zero rate corresponding to time t

Output:
B = bond price

B = 0
for i = 1 : n
   disc(i) = exp(−t_cash_flow(i) r_zero(t_cash_flow(i)) )
   B = B + v_cash_flow(i) disc(i)
end
```

using numerical integration methods such as Simpson's rule; see section 2.4. This is done by setting a tolerance $tol(i)$ for the numerical approximation of I_i and doubling the number of intervals in the partition of $[0, t_i]$ until two consecutive approximations of I_i are within $tol(i)$ of each other; see the pseudocode from Table 2.6 for more details.

From a practical standpoint, we note that the cash flow at maturity, $c_n = 100(1 + \frac{C}{m})$, is about two orders of magnitude higher than any other cash flow $c_i = 100\frac{C}{m}$, $i < n$, where m is the frequency of annual cash flows. (For example, $m = 2$ for a semiannual coupon bond.) Therefore, an optimal vector of tolerances tol has the first $n-1$ entries equal to each other, and the n-th entry two orders of magnitude smaller than the previous entry, i.e.,

$$tol = \left[\tau \ \tau \ldots \tau \ \frac{\tau}{100}\right].$$

Example: Consider a semiannual coupon bond with coupon rate 6% and maturity 20 months. Assume that the face value of the bond is 100, and that interest is compounded continuously.

(i) Compute the price of the bond if the zero rate is

$$r(0,t) = 0.0525 + \frac{\ln(1+2t)}{200}; \tag{2.66}$$

(ii) Compute the price of the bond if the instantaneous interest rate curve is

$$r(t) = 0.0525 + \frac{1}{100(1+e^{-t^2})}; \tag{2.67}$$

Table 2.6: Pseudocode for computing the bond price given the instantaneous interest rate curve

```
Input:
n = number of cash flows
t_cash_flow = vector of cash flow dates (of size n)
v_cash_flow = vector of cash flows (of size n)
r_inst(t) = instantaneous interest rate at time t
tol = vector of tolerances in the numerical approximation
      of discount factor integrals (of size n)

Output:
B = bond price

B = 0
for i = 1 : n
   I_numerical(i) = result of the numerical integration of r_inst(t)
        on the interval [0, t_cash_flow(i)] with tolerance tol(i);
   disc(i) = exp(−I_numerical(i))
   B = B + v_cash_flow(i) disc(i)
end
```

(iii) Compute the price of the bond if

$$r(t) = 0.0525 + \frac{\ln(1+2t)}{200} + \frac{t}{100(1+2t)}; \tag{2.68}$$

Answer: The bond will provide cash flows every six months, the last one being at maturity, i.e., in 20 months. Counting 6 months backward in time from maturity, we find that there are four cash flow dates until maturity, in 2 months, 8 months, 14 months, and 20 months. (Note that time is measured in years, i.e., time equal to 2 months corresponds to time equal to $\frac{2}{12}$ in the bond pricing formulas.) The coupon payments are each equal to $100 \cdot \frac{0.06}{2} = 3$. The cash flow at maturity is equal to one coupon payment plus the face value of the bond, i.e., to 103.

The input for the bond pricing codes is

$$n = 4; \quad \text{t\_cash\_flow} = \left[\frac{2}{12} \ \frac{8}{12} \ \frac{14}{12} \ \frac{20}{12}\right]; \quad \text{v\_cash\_flow} = [3 \ 3 \ 3 \ 103].$$

(i) If the zero rate $r(0,t)$ is given by (2.66), the algorithm from Table 2.5 returns $B = 101.888216$ as the price of the bond. The discount factors are

$$disc = [0.99105055 \ \ 0.96288208 \ \ 0.93400529 \ \ 0.90509128]. \tag{2.69}$$

(ii) If the instantaneous interest rate $r(t)$ is given by (2.67), the integrals

$$\text{I\_numerical}(i) = \int_0^{t\_cash\_flow(i)} r(\tau)\, d\tau, \quad \forall\, i = 1:4, \tag{2.70}$$

cannot be computed exactly, since $\int r(\tau)\, d\tau$ does not have a closed formula. We use Simpson's rule and the tolerance vector $tol = [10^{-4}\ 10^{-4}\ 10^{-4}\ 10^{-6}]$ to estimate the integrals (2.70). The bond pricing algorithm from Table 2.6 returns $B = 101.954564$ as the price of the bond.

(iii) Note that the instantaneous rate from (2.68) and the zero rate from (2.66) can be obtained from each other by using (2.40) or (2.41), and therefore the prices of the bonds from (i) and (iii) should be the same (or very close, given the numerical integration errors).

We use the bond pricing algorithm from Table 2.6 and Simpson's rule to price this bond. We choose two different tolerance vectors,

$$tol_1 = [10^{-4}\ 10^{-4}\ 10^{-4}\ 10^{-6}] \quad \text{and} \quad tol_2 = [10^{-4}\ 10^{-4}\ 10^{-4}\ 10^{-4}].$$

The bond pricing algorithm from Table 2.6 returns the following bond prices: $B_1 = 101.888217$ and $B_2 = 101.888234$, respectively.

Note that B_2 is a worse approximation than B_1 of the bond price $B = 101.888216$ obtained at (i). This is due to the fact that, when choosing the tolerance vector tol_2, we do not account for the fact that the cash flow at maturity is two orders of magnitude higher than any other cash flow. The discount factors

$$\begin{aligned} disc_1 &= [0.99105055\ \ 0.96288208\ \ 0.93400534\ \ 0.90509129] \quad \text{and} \\ disc_2 &= [0.99105055\ \ 0.96288208\ \ 0.93400534\ \ 0.90509146] \end{aligned}$$

corresponding to tol_1 and tol_2, respectively, are different on the position corresponding to the maturity of the bond, with 0.90509128 being a better approximation of the discount factor than 0.90509146; see also (2.69). □

If the yield of the bond is given, no knowledge of interest rates is needed to compute the price, the duration, and the convexity of the bond, since

$$B = \sum_{i=1}^{n} c_i e^{-yt_i}; \quad D = \frac{\sum_{i=1}^{n} t_i c_i e^{-yt_i}}{B}; \quad C = \frac{\sum_{i=1}^{n} t_i^2 c_i e^{-yt_i}}{B};$$

cf. (2.56), (2.57), and (2.61). See the pseudocode from Table 2.7 for the implementation of these formulas.

Example: Consider a semiannual coupon bond with face value 100, coupon rate 6% and maturity 20 months. If the yield of the bond is 6.50%, compute the price, the duration, and the convexity of the bond.

Table 2.7: Pseudocode for computing the price, duration and convexity of a bond given the yield of the bond

```
T = bond maturity
n = number of cash flows
t_cash_flow = vector of cash flow dates (of size n)
v_cash_flow = vector of cash flows (of size n)
y = yield of the bond

Output:
B = price of the bond
D = duration of the bond
C = convexity of the bond

B = 0; D = 0; C = 0
for i = 1 : n
   disc(i) = exp(−t_cash_flow(i) y )
   B = B + v_cash_flow(i) disc(i)
   D = D + t_cash_flow(i) v_cash_flow(i) disc(i)
   C = C + t_cash_flow(i)^2 v_cash_flow(i) disc(i)
end
D = D/B; C = C/B
```

Answer: The input for the code from Table 2.7 is

$$n = 4; \ \ y = 0.06; \ \ \text{t\_cash\_flow} = \left[\frac{2}{12}\ \frac{8}{12}\ \frac{14}{12}\ \frac{20}{12}\right]; \ \ \text{v\_cash\_flow} = [3\ 3\ 3\ 103].$$

The output of the algorithm is: bond price $B = 101.046193$, bond duration $D = 1.5804216$, and bond convexity $C = 2.5916859$. □

We note that, to compute the yield y of a bond if the price B of the bond is given, the nonlinear equation (2.56) which involves discount factors, and therefore interest rates, must be solved for y. This can be done using, e.g., Newton's method; see section 8.5 for more details.

2.9 References

A comprehensive presentation of numerical integration methods can be found in the numerical analysis monograph Stoer and Bulirsch [28]. A very interesting treatment of interest rate curves, including the LIBOR model, is given in Neftci [19].

2.10 Exercises

1. Compute the integral of the function $f(x,y) = x^2 - 2y$ on the region bounded by the parabola $y = (x+1)^2$ and the line $y = 5x - 1$.

2. Let $f : (0, \infty) \to \mathbb{R}$ denote the Gamma function, i.e., let

$$f(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx.$$

(i) Show that $f(\alpha)$ is well defined for any $\alpha > 0$, i.e., show that both

$$\int_0^1 x^{\alpha-1} e^{-x} \, dx = \lim_{t \to 0} \int_t^1 x^{\alpha-1} e^{-x} \, dx$$

and

$$\int_1^\infty x^{\alpha-1} e^{-x} \, dx = \lim_{t \to \infty} \int_1^t x^{\alpha-1} e^{-x} \, dx$$

exist and are finite.

(ii) Prove, using integration by parts, that $f(\alpha) = (\alpha - 1) f(\alpha - 1)$ for any $\alpha > 1$. Show that $f(1) = 1$ and conclude that, for any $n \geq 1$ positive integer, $f(n) = (n-1)!$.

3. Compute an approximate value of $\int_1^3 \sqrt{x}\, e^{-x} dx$ using the Midpoint rule, the Trapezoidal rule, and Simpson's rule. Start with $n = 4$ intervals, and double the number of intervals until two consecutive approximations are within 10^{-6} of each other.

4. Let $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{x^{5/2}}{1+x^2}$.

(i) Use Midpoint rule with $tol = 10^{-6}$ to compute an approximation of

$$I = \int_0^1 f(x) \, dx = \int_0^1 \frac{x^{5/2}}{1+x^2}. \qquad (2.71)$$

(ii) Show that $f^{(4)}(x)$ is not bounded on the interval $(0, 1)$.

(iii) Apply Simpson's rule with $n = 2^k$, $k = 2 : 8$, intervals to compute the integral I from (2.71). Conclude that Simpson's rule converges, although the theoretical estimate (2.26) cannot be used to infer convergence in this case since $f^{(4)}(x)$ is unbounded on $(0, 1)$.

5. Let K, T, σ and r be positive constants. Define the function $g : \mathbb{R} \to \mathbb{R}$ as

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b(x)} e^{-\frac{y^2}{2}} \, dy,$$

where $b(x) = \left(\ln\left(\frac{x}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) T \right) / \left(\sigma\sqrt{T}\right)$. Compute $g'(x)$.

Note: This function is related to the Delta of a plain vanilla Call option; see Section 3.5 for more details.

6. Let $h(x)$ be a function such that $\int_{-\infty}^{\infty} |xh(x)| dx$ exists. Define $g(t)$ by

$$g(t) = \int_t^\infty (x - t) h(x) \, dx,$$

and show that

$$\frac{\partial^2 g}{\partial t^2} = h(t).$$

Note: The price of a call option can be regarded as a function of the strike price K. By using risk–neutral valuation (see section 4.7), we find that

$$\begin{aligned} C(K) &= e^{-rT} E_{RN}[\max(S(T) - K, 0)] \\ &= e^{-rT} \int_{-\infty}^\infty \max(x - K, 0) f(x) \, dx \\ &= e^{-rT} \int_K^\infty (x - K) f(x) \, dx \\ &= \int_K^\infty (x - K) h(x) \, dx, \end{aligned}$$

where $f(x)$ is the probability density function of $S(T)$ given $S(0)$, and $h(x) = e^{-rT} f(x)$. Then, according to the result of this exercise,

$$\frac{\partial^2 C}{\partial K^2} = e^{-rT} f(K).$$

7. The continuously compounded 6-month, 12-month, 18-month, and 24-month zero rates are 5%, 5.25%, 5.35%, and 5.5%, respectively. Find the price of a two year semiannual coupon bond with coupon rate 5%.

8. The continuously compounded 6-month, 12-month, 18-month, and 24-month zero rates are 5%, 5.25%, 5.35%, and 5.5%, respectively. What is the par yield for a 2-year semiannual coupon bond?

9. Assume that the continuously compounded instantaneous interest curve has the form

$$r(t) = 0.05 + 0.005 \ln(1+t), \quad \forall\, t \geq 0.$$

(i) Find the corresponding zero rate curve;

(ii) Compute the 6-month, 12-month, 18-month, and 24-month discount factors;

(iii) Find the price of a two year semiannual coupon bond with coupon rate 5%.

10. The yield of a semiannual coupon bond with 6% coupon rate and 30 months to maturity is 9%. What are the price, duration and convexity of the bond?

11. The yield of a 14 months quarterly coupon bond with 8% coupon rate is 7%. Compute the price, duration, and convexity of the bond.

12. Compute the price, duration and convexity of a two year semiannual coupon bond with face value 100 and coupon rate 8%, if the zero rate curve is given by $r(0,t) = 0.05 + 0.01 \ln\left(1 + \frac{t}{2}\right)$.

13. If the coupon rate of a bond goes up, what can be said about the value of the bond and its duration? Give a financial argument. Check your answer mathematically, i.e., by computing $\frac{\partial B}{\partial C}$ and $\frac{\partial D}{\partial C}$, and showing that these functions are either always positive or always negative.

14. By how much would the price of a ten year zero-coupon bond change if the yield increases by ten basis points? (One percentage point is equal to 100 basis points. Thus, 10 basis points is equal to 0.001.)

15. A five year bond with duration $3\frac{1}{2}$ years is worth 102. Find an approximate price of the bond if the yield decreases by fifty basis points.

16. Establish the following relationship between duration and convexity:

$$C = D^2 - \frac{\partial D}{\partial y}$$

Chapter 3

Probability concepts. Black–Scholes formula. Greeks and Hedging.

Discrete probability concepts.

Continuous probability concepts. Random variables. Probability density and cumulative distribution. Mean, variance, covariance and correlation.

Normal random variables and the standard normal variable.

3.1 Discrete probability concepts

Let $S = \{s_1, s_2, \ldots, s_n\}$ be a finite set and let $P : S \to [0,1]$ be a probability functiondefined on S, i.e., a function with the following properties:

$$P(s_i) \geq 0, \ \forall\, i = 1:n, \quad \text{and} \quad \sum_{i=1}^{n} P(s_i) = 1. \tag{3.1}$$

Any function $X : S \to \mathbb{R}$ is called a random variable defined on the set of outcomes S.

Definition 3.1. *Let $X : S \to \mathbb{R}$ be a random variable on the set S endowed with a probability function $P : S \to [0,1]$.*
The expected value $E[X]$ of X (also called the mean of X) is

$$E[X] = \sum_{i=1}^{n} P(s_i) X(s_i). \tag{3.2}$$

The variance $var(X)$ of X is

$$var(X) = E[(X - E[X])^2]. \tag{3.3}$$

The standard deviation $\sigma(X)$ of X is

$$\sigma(X) = \sqrt{var(X)}. \tag{3.4}$$

Lemma 3.1. *Let $X : S \to \mathbb{R}$ be a random variable on the set S endowed with a probability function $P : S \to [0,1]$. Then,*

$$var(X) = E[X^2] - (E[X])^2 = \sum_{i=1}^{n} P(s_i)\,(X(s_i))^2 - (E[X])^2.$$

Proof. Let $m = E[X]$. From (3.3), and using (3.1) and (3.2), it follows that

$$\begin{aligned} var[X] &= E[(X - E[X])^2] \\ &= \sum_{i=1}^{n} P(s_i)\,(X(s_i) - m)^2 \\ &= \sum_{i=1}^{n} P(s_i)\,\left(X^2(s_i) - 2mX(s_i) + m^2\right) \\ &= \sum_{i=1}^{n} P(s_i)X^2(s_i) \;-\; 2m\sum_{i=1}^{n} P(s_i)X(s_i) \;+\; m^2\sum_{i=1}^{n} P(s_i) \\ &= E[X^2] \;-\; 2m \cdot m \;+\; m^2 \\ &= E[X^2] \;-\; m^2 \\ &= E[X^2] \;-\; (E[X])^2. \end{aligned}$$

□

Example: You throw a pair of fair dice. You win \$10 if the sum of the dice is 10, or lose \$1 otherwise. Is the game fair?

Answer: Let x and y be the outcomes of the throw corresponding to the first dice and to the second dice, respectively. There are 36 equally probable possible outcomes (x,y) for the game, of probability $1/36$ each, corresponding to $x = 1:6$ and $y = 1:6$.

In our discrete probability setting, let

$$S = \{\ (x,y) \text{ with } x \in \{1,2,\ldots,6\} \text{ and } y \in \{1,2,\ldots,6\}\ \}.$$

Let $P : S \to [0,\ 1]$ be the probability function on S given by

$$P(x,y) = \frac{1}{36}, \quad \forall\ (x,y) \in S.$$

The value of the winnings for each outcome is, by definition, a random variable on the set of outcomes S, which we denote by $X : S \to [0,\ 1]$. Thus,

$$X(x,y) = \begin{cases} 10, & \text{if } x + y = 10; \\ -1, & \text{otherwise.} \end{cases}$$

The event that $x + y = 10$ happens for only three outcomes: $(6,4)$, $(4,6)$, and $(5,5)$. The payoff in either one of these cases is \$10. For all the other 33 outcomes \$1 is lost, since $x + y \neq 10$.

The expected value of your winnings, i.e., the expected value of X, is what dictates whether the game is fair (if the expected value is zero) or not (if the expected value is positive or negative). From (3.2), we find that

$$E[X] = \sum_{(x,y)\in S} P(x,y)X(x,y) = 3\,\frac{1}{36}\,10 \;+\; 33\,\frac{1}{36}\,(-1) = -\frac{1}{12}.$$

The game is not fair. □

Example: A stock with spot price S_0 at time 0 has the following model for its evolution: over each of three consecutive time intervals of length τ beginning at time 0, the price of the stock will either go up by a factor $u > 1$ with probability p, or down by a factor $d < 1$ with probability $1 - p$. (Successive moves of the price of the stock are assumed to be independent.) What is the probability space corresponding to the value S_T of the stock at time $T = 3\tau$?

Answer: The probability space S is the set of all different paths that the stock could follow three consecutive time intervals, i.e.,

$$S = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\},$$

where U represents an "up" move and D represents a "down" move.

The probability function $P : S \to [0,\ 1]$ gives the probability that the stock follows an individual path from time 0 to time T. For example,

$$\begin{aligned} P(UDD) &= P(U)P(D)P(D) = p(1-p)^2; \\ P(DUU) &= P(D)P(U)P(U) = p^2(1-p). \end{aligned}$$

The value S_T of the stock at time T is a random variable defined on S, i.e., $S_T : S \to [0,\ \infty)$. For example,

$$\begin{aligned} S_T(UDD) &= S_0 u d^2; \\ S_T(DUU) &= S_0 u^2 d. \end{aligned}$$ □

3.2 Continuous probability concepts

Let S be a set, and consider a family $\mathbb{F}$ of subsets of S such that
(i) $S \in \mathbb{F}$;
(ii) if $A \in \mathbb{F}$, then $S \setminus A \in \mathbb{F}$;

(iii) if $A_k \in \mathbb{F}$, $k = 1 : \infty$, then $\bigcup_{k=1}^{\infty} A_k \in \mathbb{F}$.

The set S represents the sample space, any element of S is a possible outcome, and a set $A \in \mathbb{F}$ is an event[1], i.e., a collection of possible outcomes.

A probability function $P : \mathbb{F} \to [0, 1]$ defined on $\mathbb{F}$ is a nonnegative function such that $P(S) = 1$ and

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k), \quad \forall\ A_k \in \mathbb{F},\ k = 1 : \infty, \quad \forall\ A_i \cap A_j = \emptyset,\ i \neq j.$$

A function $X : S \to \mathbb{R}$ is called a random variable if

$$\{\ s \in S \text{ such that } X(s) \leq t\ \} \ \in\ \mathbb{F}, \quad \forall\ t \in \mathbb{R}.$$

The cumulative distribution function $F : \mathbb{R} \to [0,\ 1]$ of the random variable X is defined as

$$F(t) \ = \ P(X \leq t). \tag{3.5}$$

For any random variable X considered here, we assume[2] that X has a probability density function $f(x)$, i.e., we assume that there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$P(a \leq X \leq b) \ = \ \int_a^b f(x)dx. \tag{3.6}$$

For an integrable function $f(x)$ to be the probability density function of a random variable it is necessary and sufficient that $f(x) \geq 0$, for all $x \in \mathbb{R}$, and

$$\int_{-\infty}^{\infty} f(x)\ dx \ = \ 1. \tag{3.7}$$

Lemma 3.2. *Let $f(x)$ be the probability density function of the random variable X. Then the cumulative distribution function of X can be written as*

$$F(t) \ = \ \int_{-\infty}^{t} f(x)\ dx. \tag{3.8}$$

Definition 3.2. *Let $f(x)$ be the probability density function of the random variable X. The expected value $E[X]$ of X (also called the mean of X) is defined as*

$$E[X] \ = \ \int_{-\infty}^{\infty} x f(x)\ dx. \tag{3.9}$$

[1]A family of subsets $\mathbb{F}$ of S satisfying the properties (i)–(iii) is called a sigma-algebra of S. We introduce the concepts of continuous probability without formally discussing measure theory topics such as sigma-algebras and measurable functions. For more details on measure theory, we refer the reader to Royden [22] or Rudin [23].

[2]It can be shown that the random variable X has a probability density function if and only if the cumulative density function $F(t)$ of X is an absolutely continuous function.

The expected values of a random variables is a linear operator:

Lemma 3.3. *Let X and Y be two random variables over the same probability space. Then*

$$E[X+Y] \ = \ E[X] \ + \ E[Y]; \tag{3.10}$$

$$E[cX] \ = \ c\ E[X], \quad \forall\ c \in \mathbb{R}. \tag{3.11}$$

The following result will be used throughout this chapter, and is presented without proof:

Lemma 3.4. *Let $h : \mathbb{R} \to \mathbb{R}$ be a piecewise continuous function, and let $X : S \to \mathbb{R}$ be a random variable with probability density function $f(x)$ such that $\int_{\mathbb{R}} |h(x) f(x)\ dx < \infty$. Then $h(X) : S \to \mathbb{R}$ is a random variable and the expected value of $h(X)$ is*

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)\ f(x)\ dx. \tag{3.12}$$

3.2.1 Variance, covariance, and correlation

The variance and standard deviation of a random variable offer important information about its distribution.

Definition 3.3. *Let $f(x)$ be the probability density function of the random variable X, and assume that $\int_{\mathbb{R}} x^2 f(x)\ dx < \infty$. Let $m = E[X]$ denote the expected value of X. The variance $var(X)$ of X is defined as*

$$var(X) \ = \ E[(X - E[X])^2] \ = \ \int_{-\infty}^{\infty} (x - m)^2\ f(x)\ dx. \tag{3.13}$$

The standard deviation $\sigma(X)$ of X is defined as

$$\sigma(X) \ = \ \sqrt{var(X)}. \tag{3.14}$$

Therefore, $var(X) \ = \ (\sigma(X))^2 \ = \ \sigma^2(X)$.

From (3.13) and (3.14), it is easy to see that

$$\text{var}(cX) \ = \ c^2\ \text{var}(X), \quad \forall\ c \in \mathbb{R}; \tag{3.15}$$

$$\sigma(cX) \ = \ |c|\ \sigma(X), \quad \forall\ c \in \mathbb{R}. \tag{3.16}$$

The next result corresponds to that of Lemma 3.1, and its proof follows along the same lines:

Lemma 3.5. *If X is a random variable, then*

$$var(X) = E[X^2] - (E[X])^2.$$

Definition 3.4. *Let X and Y be two random variables over the same probability space. The covariance $cov(X,Y)$ of X and Y is defined as*

$$cov(X,Y) = E[(X - E[X])\,(Y - E[Y])]. \tag{3.17}$$

The correlation $corr(X,Y)$ between X and Y is equal to the covariance of X and Y normalized with respect to the standard deviations of X and Y, i.e.,

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma(X)\,\sigma(Y)}, \tag{3.18}$$

where $\sigma(X)$ and $\sigma(Y)$ are the standard deviations of X and Y, respectively.

Lemma 3.6. *Let X and Y be two random variables over the same probability space. Then*

$$cov(X,Y) = E[XY] - E[X]E[Y]. \tag{3.19}$$

Proof. Using (3.10) and (3.11) repeatedly, we find that

$$\begin{aligned} \text{cov}(X,Y) &= E[\,(X - E[X])\,(Y - E[Y])\,] \\ &= E[\,XY - XE[Y] - YE[X] + E[X]E[Y]\,] \\ &= E[XY] - E[\,XE[Y]\,] - E[\,YE[X]\,] + E[X]E[Y]. \end{aligned} \tag{3.20}$$

Since $E[X]$ and $E[Y]$ are constants, we conclude from (3.11) that

$$E[\,XE[Y]\,] = E[Y]\,E[X] \quad \text{and} \quad E[\,YE[X]\,] = E[X]\,E[Y].$$

Therefore, formula (3.20) becomes

$$\text{cov}(X,Y) = E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y].$$

□

Let X, Y, and U be random variables over the same probability space. The following properties are easy to establish and show that the covariance of two random variables is a symmetric bilinear operator:

$$\begin{aligned} \text{cov}(X,Y) &= \text{cov}(Y,X); \\ \text{cov}(X+U,Y) &= \text{cov}(X,Y) + \text{cov}(U,Y); \\ \text{cov}(X,Y+U) &= \text{cov}(X,Y) + \text{cov}(X,U); \\ \text{cov}(cX,Y) &= \text{cov}(X,cY) = c\,\text{cov}(X,Y), \ \forall\, c \in \mathbb{R}; \quad (3.21) \\ \text{cov}(c_1X,c_2Y) &= c_1c_2\,\text{cov}(X,Y), \ \forall\, c_1,c_2 \in \mathbb{R}. \quad (3.22) \end{aligned}$$

Lemma 3.7. *Let X and Y be two random variables over the same probability space. Then,*

$$var(X+Y) = var(X) + 2cov(X,Y) + var(Y), \tag{3.23}$$

or, equivalently,

$$var(X+Y) = \sigma^2(X) + 2\sigma(X)\,\sigma(Y)\,corr(X,Y) + \sigma^2(Y), \tag{3.24}$$

where $\sigma(X)$ and $\sigma(Y)$ are the standard deviation of X and Y, respectively.

Proof. Formula (3.23) is derived from definitions (3.13) and (3.17), by using the additivity of the expected value (3.10), as follows:

$$\begin{aligned} var(X+Y) &= E\Big[\,(\,(X+Y) - E[X+Y]\,)^2\,\Big] \\ &= E\Big[\,(\,(X - E[X]) + (Y - E[Y])\,)^2\,\Big] \\ &= E\,[\,(X - E[X])^2\,] + 2E[\,(X - E[X])\,(Y - E[Y])\,] \\ &\quad + E\,[\,(Y - E[Y])^2\,] \\ &= \text{var}(X) + 2\text{cov}(X,Y) + \text{var}(Y). \end{aligned}$$

Formula (3.24) is a direct consequence of (3.18) and (3.23). □

The relevant information contained by the covariance of two random variables is related to whether they are positively or negatively correlated, i.e., whether $\text{cov}(X,Y) > 0$ or $\text{cov}(X,Y) < 0$. The correlation of the two variables contains the same information as the covariance, in terms of its sign, but its size is also relevant, since it is scaled to adjust for multiplication by constants, i.e.,

$$\text{corr}(c_1X,c_2Y) = \text{sign}(c_1c_2)\,\text{corr}(X,Y), \ \forall\, c_1,c_2 \in \mathbb{R}. \tag{3.25}$$

Here, the sign function $\text{sgn}(c)$ is equal to 1, if $c > 0$, and to -1 is $c < 0$.

To see this, recall from (3.16) that $\sigma(c_1X) = |c_1|\sigma(X)$ and $\sigma(c_2Y) = |c_2|\sigma(Y)$ Then, from (3.18) and using (3.22), it follows that

$$\begin{aligned} \text{corr}(c_1X,c_2Y) &= \frac{\text{cov}(c_1X,c_2Y)}{\sigma(c_1X)\,\sigma(c_2Y)} = \frac{c_1c_2\,\text{cov}(X,Y)}{|c_1|\,|c_2|\,\sigma(X)\,\sigma(Y)} \\ &= \frac{c_1c_2}{|c_1c_2|}\,\frac{\text{cov}(X,Y)}{\sigma(X)\,\sigma(Y)} = \text{sgn}(c_1c_2)\,\text{corr}(X,Y). \end{aligned}$$

Lemma 3.8. *Let X and Y be two random variables over the same probability space. Then*

$$-1 \le corr(X,Y) \le 1. \tag{3.26}$$

Proof. Let $\alpha \in \mathbb{R}$ be an arbitrary constant. From (3.24), (3.21), (3.15), and (3.18) we find that

$$\begin{aligned} var(X+\alpha Y) &= \mathrm{var}(X) + 2\mathrm{cov}(X,\alpha Y) + \mathrm{var}(\alpha Y) \\ &= \mathrm{var}(X) + 2\alpha\ \mathrm{cov}(X,Y) + \alpha^2 \mathrm{var}(Y) \\ &= \sigma^2(X) + 2\alpha\ \sigma(X)\ \sigma(Y)\ \mathrm{corr}(X,Y) + \alpha^2\sigma^2(Y). \end{aligned}$$

Since $var(X+\alpha Y) \geq 0$, we obtain that

$$\alpha^2\ \sigma^2(Y) + 2\alpha\ \sigma(X)\sigma(Y)\mathrm{corr}(X,Y) + \sigma^2(X) \ \geq\ 0, \quad \forall\ \alpha \in \mathbb{R}. \tag{3.27}$$

Note that the left hand side of (3.27) is a quadratic polynomial of α. The inequality (3.27) holds true for any real number α if and only if this polynomial has at most one real double root, i.e., if its discriminant is nonpositive:

$$(2\sigma(X)\sigma(Y)\mathrm{corr}(X,Y))^2 - 4\sigma^2(X)\sigma^2(Y) \ \leq\ 0.$$

This is equivalent to $4\sigma^2(X)\sigma^2(Y)(\mathrm{corr}(X,Y))^2 \leq 4\sigma^2(X)\sigma^2(Y)$, which happens if and only if $|\mathrm{corr}(X,Y)| \leq 1$, which is equivalent to (3.26). □

The elegant technique used in the proof of Lemma 3.8 can also be applied, e.g., for showing that the inner product of two vectors is bounded from above by the product of the norms of the vectors[3]:

$$\left(\sum_{i=1}^n x_i\ y_i\right)^2 \ \leq\ \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right), \quad \forall\ x_i, y_i \in \mathbb{R},\ i = 1:n. \tag{3.28}$$

The expected value and variance of a linear combination of n random variables can be obtained by induction from the corresponding formulas (3.10–3.11) and (3.23–3.24) for two random variables.

Lemma 3.9. *Let X_i, $i=1:n$, be random variables over the same probability space, and let $c_i \in \mathbb{R}$ be real numbers. Then,*

$$E\left[\sum_{i=1}^n c_i X_i\right] = \sum_{i=1}^n c_i E[X_i]; \tag{3.29}$$

$$var\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 var(X_i) + 2 \sum_{1\leq i<j\leq n} c_i c_j cov(X_i, X_j). \tag{3.30}$$

[3]A proof for (3.28), also called the Cauchy–Schwarz inequality, can be given using the fact that

$$\sum_{i=1}^n (x_i + \alpha y_i)^2 \ \geq\ 0, \quad \forall\ \alpha \in \mathbb{R}.$$

If $\rho_{i,j} = corr(X_i, X_j)$, $1 \leq i < j \leq n$, is the correlation between X_i and X_j, and if $\sigma_i^2 = var(X_i)$, $i = 1:n$, then (3.30) can also be written as

$$var\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2\sigma_i^2 + 2 \sum_{1\leq i<j\leq n} c_i c_j \sigma_i \sigma_j corr(X_i, X_j). \tag{3.31}$$

3.3 The standard normal variable

The standard normal variable, denoted by Z, is the random variable with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}. \tag{3.32}$$

It is easy to see that $f(x) > 0$, $\forall\ x \in \mathbb{R}$. To conclude that $f(x)$ is indeed a density function, we would have to show that

$$\int_{-\infty}^{\infty} f(x)\ dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx = 1. \tag{3.33}$$

We postpone the proof of this fact until section 7.5 and assume for now that (3.33) holds true.

The following result will be needed in the proof of Lemma 3.10:

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\ dx = \sqrt{2\pi}. \tag{3.34}$$

Note that $x^2 e^{-\frac{x^2}{2}}$ is an even function. From Lemma 0.1, it follows that

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\ dx = 2\int_0^{\infty} x^2 e^{-\frac{x^2}{2}}\ dx = 2 \lim_{t\to\infty} \int_0^t x^2 e^{-\frac{x^2}{2}}\ dx. \tag{3.35}$$

Note that $\left(-e^{-\frac{x^2}{2}}\right)' = xe^{-\frac{x^2}{2}}$. By integration by parts, we find that

$$\begin{aligned} \int_0^t x^2 e^{-\frac{x^2}{2}}\ dx &= \int_0^t x\left(-e^{-\frac{x^2}{2}}\right)'\ dx \\ &= \left(-xe^{-\frac{x^2}{2}}\right)\Big|_0^t - \int_0^t \left(-e^{-\frac{x^2}{2}}\right)\ dx \\ &= -te^{-\frac{t^2}{2}} + \int_0^t e^{-\frac{x^2}{2}}\ dx. \end{aligned} \tag{3.36}$$

From (3.35) and (3.36) it follows that

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = 2 \lim_{t\to\infty} \left(-te^{-\frac{t^2}{2}} + \int_0^t e^{-\frac{x^2}{2}}\, dx \right)$$
$$= 2 \lim_{t\to\infty} \int_0^t e^{-\frac{x^2}{2}}\, dx = 2\int_0^\infty e^{-\frac{x^2}{2}}\, dx. \qquad (3.37)$$

Recall from (3.33), that $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$. Since $e^{-\frac{x^2}{2}}$ is an even function, we obtain from Lemma 0.1 that

$$\int_0^\infty e^{-\frac{x^2}{2}}\, dx = \frac{1}{2}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \frac{1}{2}\sqrt{2\pi} = \sqrt{\frac{\pi}{2}}. \qquad (3.38)$$

From (3.37) and (3.38), it follows that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}}\cdot 2\int_0^\infty e^{-\frac{x^2}{2}}\, dx = 1,$$

and therefore (3.34) holds true.

Lemma 3.10. *The standard normal variable has mean 0 and variance 1, i.e.,*

$$E[Z] = 0; \qquad (3.39)$$
$$var(Z) = 1. \qquad (3.40)$$

Proof. By definition (3.9), and using the fact that $xe^{-\frac{x^2}{2}}$ is an odd function and Lemma 0.2, we obtain that

$$E[Z] = \int_{-\infty}^{\infty} xf(x)\, dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}\, dx = 0.$$

To compute the variance of Z, note that

$$\mathrm{var}(Z) = E[(Z - E[Z])^2] = E[Z^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = 1,$$

where the last equality follows from (3.34). □

Lemma 3.11. *If Z is the standard normal variable, then $E[Z^2] = 1$.*

Proof. From Lemma 3.5, it follows that $\mathrm{var}(Z) = E[Z^2] - (E[Z])^2$. Since $\mathrm{var}(Z) = 1$ and $E[Z] = 0$, see (3.39–3.40), we conclude that $E[Z^2] = 1$. □

Definition 3.5. *Denote by $N(t)$ the cumulative distribution of the standard normal variable Z. Then,*

$$N(t) = P(Z \le t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx. \qquad (3.41)$$

Lemma 3.12. *Let Z be the standard normal variable. Then,*

$$P(Z \ge a) = P(Z \le -a), \quad \forall\ a \in \mathbb{R}. \qquad (3.42)$$

In other words, if $N(t)$ is the cumulative distribution of Z, then

$$1 - N(a) = N(-a), \quad \forall\ a \in \mathbb{R}. \qquad (3.43)$$

Proof. By definition (3.6), and using (3.32), we find that

$$P(Z \ge a) = \frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{x^2}{2}}\, dx.$$

We use the substitution $x = -y$. Then $dx = -dy$ and the limits of integration change from $x = a$ to $y = -a$ and from $x = \infty$ to $y = -\infty$. Thus,

$$P(Z \ge a) = \frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}}\int_{-a}^{-\infty} e^{-\frac{(-y)^2}{2}}\, (-dy)$$
$$= -\frac{1}{\sqrt{2\pi}}\int_{-a}^{-\infty} e^{-\frac{y^2}{2}}\, dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-a} e^{-\frac{y^2}{2}}\, dy$$
$$= P(Z \le -a),$$

and (3.42) is proven.

To obtain (3.43) from (3.42) note that, by definition (3.41),

$$P(Z \ge a) = 1 - P(Z \le a) = 1 - N(a) \quad \text{and} \quad P(Z \le -a) = N(-a).$$

□

3.4 Normal random variables

Definition 3.6. *The random variable X is a normal variable if and only if*

$$X = \mu + \sigma Z, \qquad (3.44)$$

where Z is the standard normal variable and $\mu, \sigma \in \mathbb{R}$.

Lemma 3.13. *Let $X = \mu + \sigma Z$ be a normal variable. Then*

$$E[X] = \mu; \qquad (3.45)$$
$$var(X) = \sigma^2; \qquad (3.46)$$
$$\sigma(X) = |\sigma| \qquad (3.47)$$

Proof. Using Lemma 3.3 and the fact that $E[Z] = 0$, it follows that

$$E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu.$$

From (3.13), and using Lemma 3.3 and the fact that $E[Z^2] = 1$, cf. Lemma 3.11, we find that

$$\mathrm{var}(X) = E[(X - E[X])^2] = E[(\mu + \sigma Z - \mu)^2] = E[\sigma^2 Z^2] = \sigma^2 E[Z^2] = \sigma^2.$$

□

Note that the constant random variable $X = \mu$ is a degenerate normal variable with mean μ and standard deviation 0.

Lemma 3.14. *Let $X = \mu + \sigma Z$ be a normal variable, with $\sigma \neq 0$, and denote by $h(x)$ the probability density function of X. Then,*

$$h(x) = \frac{1}{|\sigma|\sqrt{2\pi}} \, exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \qquad (3.48)$$

Proof. From definition (3.6), we know that

$$P(a \leq X \leq b) = \int_a^b h(y)\, dy, \quad \forall \, a < b. \qquad (3.49)$$

Assume that $\sigma > 0$. Since $X = \mu + \sigma Z$, we find that

$$\begin{aligned} P(a \leq X \leq b) &= P(a \leq \mu + \sigma Z \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \exp\left(-\frac{x^2}{2}\right) dx, \qquad (3.50) \end{aligned}$$

since $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ is the density function of Z.

We make the substitution $x = \frac{y-\mu}{\sigma}$ in (3.50). The integration limits change from $x = \frac{a-\mu}{\sigma}$ to $y = a$ and from $x = \frac{b-\mu}{\sigma}$ to $y = b$. Also, $dx = \frac{dy}{\sigma}$. Therefore, we find from (3.50) that

$$P(a \leq X \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy, \quad \forall \, a < b. \qquad (3.51)$$

From (3.49) and (3.51), it follows that the density function $h(x)$ of X is

$$h(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = \frac{1}{|\sigma|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

since $\sigma > 0$.

If $\sigma < 0$, then (3.50) becomes

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \geq Z \geq \frac{b-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{(b-\mu)/\sigma}^{(a-\mu)/\sigma} \exp\left(-\frac{x^2}{2}\right) dx.$$

The substitution $x = \frac{y-\mu}{\sigma}$ yields

$$\begin{aligned} P(a \leq X \leq b) &= \frac{1}{\sigma\sqrt{2\pi}} \int_b^a \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= -\frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy, \end{aligned}$$

and therefore

$$h(y) = -\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = \frac{1}{|\sigma|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

since $|\sigma| = -\sigma$ if $\sigma < 0$. □

Example: Let $X = \mu + \sigma Z$ be a normal variable, and let $Y = 2\mu - X$. Show that X and Y have the same probability densities.

Answer: Note that $Y = 2\mu - X = \mu - \sigma Z$ is a normal variable. Let $h_X(x)$ and $h_Y(x)$ be the probability density functions of X and Y, respectively. From (3.48), it follows that

$$\begin{aligned} h_Y(x) &= \frac{1}{|-\sigma|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2(-\sigma)^2}\right) \\ &= \frac{1}{|\sigma|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = h_X(x). \quad \square \end{aligned}$$

We note that the sum two *independent* normal variables (and, in general, the sum of any finite number of independent normal variables) is also a normal variable. This is not the case if the normal variables are not independent; see section 4.3 for more details.

FINANCIAL APPLICATIONS

The Black–Scholes formula.

The Greeks of plain vanilla European call and put options.

Implied volatility.

The concept of hedging. Δ–hedging and Γ–hedging for options.

Implementation of the Black–Scholes formula.

3.5 The Black–Scholes formula

The Black–Scholes formulas give the price of plain vanilla European call and put options, under the assumption that the price of the underlying asset[4] has lognormal distribution. A detailed discussion of the lognormal assumption can be found in section 4.6.

To introduce the Black–Scholes formula, it is enough to assume that, for any values of t_1 and t_2 with $t_1 < t_2$, the random variable $\frac{S(t_2)}{S(t_1)}$ is lognormal with parameters $\left(\mu - q - \frac{\sigma^2}{2}\right)(t_2 - t_1)$ and $\sigma^2 (t_2 - t_1)$, i.e.,

$$\ln\left(\frac{S(t_2)}{S(t_1)}\right) = \left(\mu - q - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma\sqrt{t_2 - t_1}Z, \tag{3.52}$$

where Z is the standard normal variable. The constants μ and σ are called the drift and the volatility of the price $S(t)$ of the underlying asset and represent the expected value and the standard deviation of the returns of the asset; q is the continuous rate at which the asset pays dividends; see section 4.6 for more details.

In the Black–Scholes formulas, the price of a plain vanilla European option depends on the following parameters:

- S, the spot price of the underlying asset at time t;
- T, the maturity of the option; note that the time to maturity is $T - t$;
- K, the strike price of the option;
- r, the risk–free interest rate, assumed to be constant over the life of the option, i.e., between t and T;
- σ, the volatility of the underlying asset, i.e., the standard deviation of the returns of the asset;
- q, the dividend rate of the underlying asset; the asset is assumed to pay dividends at a continuous rate.

[4]Examples of underlying assets are stocks, futures, indeces, etc.

The Black–Scholes Formulas for European Call and Put Options:

Assume that the price of the underlying asset has lognormal distribution and volatility σ, that the asset pays dividends continuously at the rate q, and that the risk–free interest rate is constant and equal to r. Let $C(S,t)$ be the value at time t of a call option with strike K and maturity T, and let $P(S,t)$ be the value at time t of a put option with strike K and maturity T. Then,

$$C(S,t) = Se^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2); \tag{3.53}$$

$$P(S,t) = Ke^{-r(T-t)}N(-d_2) - Se^{-q(T-t)}N(-d_1), \tag{3.54}$$

where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}; \tag{3.55}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{3.56}$$

Here, $N(z)$ is the cumulative distribution of the standard normal variable:

$$N(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}\,dx.$$

For a non-dividend-paying asset, i.e., for $q = 0$, the Black–Scholes formulas are

$$C(S,t) = SN(d_1) - Ke^{-r(T-t)}N(d_2); \tag{3.57}$$

$$P(S,t) = Ke^{-r(T-t)}N(-d_2) - SN(-d_1), \tag{3.58}$$

where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}; \tag{3.59}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{3.60}$$

Note that the general rule of thumb that, for a continuous dividend paying asset with dividend rate q, the Black–Scholes formulas (3.53) and (3.54) can be obtained from the formulas (3.57) and (3.58) for a non-dividend paying asset by substituting $r - q$ for r, is *wrong*. The reason for this is somewhat subtle: while r represents the risk–neutral drift of the underlying asset in (3.59) and (3.60), and therefore (3.55) and (3.56) can be obtained from (3.59)

and (3.60) by substituting $r - q$ for r, the term $e^{-r(T-t)}$ from (3.57) and (3.58) is the discount factor from time T to time t and does not change in the formulas (3.53) and (3.54) corresponding to the dividend–paying asset.

In general, after a result is obtained for a European call option, the corresponding result for a European put option can be derived by using the Put–Call parity. For example, the Black–Scholes formula (3.54) for the European put, can be obtained from the Black–Scholes formula (3.53) for the European call as follows: From (3.53), we know that

$$C(S,t) \;=\; Se^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2).$$

Then, from the Put-Call parity formula (1.47), we find that

$$\begin{aligned} P(S,t) &= Ke^{-r(T-t)} - Se^{-q(T-t)} + C(S,t) \\ &= Ke^{-r(T-t)} - Se^{-q(T-t)} + (Se^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2)) \\ &= Ke^{-r(T-t)}\,(1 - N(d_2)) \;-\; Se^{-q(T-t)}\,(1 - N(d_1)) \\ &= Ke^{-r(T-t)}N(-d_2) - Se^{-q(T-t)}N(-d_1), \end{aligned}$$

since $1 - N(a) = N(-a)$, for any $a \in \mathbb{R}$; cf. Lemma 3.12.

If $V(S(t), K)$ denotes the value at time t of a plain vanilla options with strike K and spot price $S(t)$ of the underlying asset, it is easy to see, from (3.53–3.56), that $V(aS(t), aK) = aV(S(t), K)$ for any $a > 0$. Therefore,

$$V(S(t),K) \;=\; K\,V\left(\frac{S(t)}{K}, 1\right) \;=\; K\,V\left(1 + \frac{S(t)-K}{K}, 1\right).$$

Thus, the relevant factor when pricing an option with strike K is

$$\frac{S-K}{K} = \frac{S(t)-K}{K},$$

which represents the percentage by which an option is in–the–money or out–of–the–money at time t.

Example: A call option with strike $K = 60$ is 15% out of the money. Find the spot price of the underlying asset.

Answer: For an out–of–money call, the spot price is smaller than the strike price, i.e., $S(t) < K$. The call is 15% out of the money if

$$\frac{K - S(t)}{K} \;=\; 0.15,$$

i.e., if $S(t) = 0.85K$. For $K = 60$, we find that $S(t) = 51$. □

3.6 The Greeks of European call and put options

Let V be the value of a portfolio of derivative securities on one underlying asset[5]. The rates of change of the value of the portfolio with respect to various parameters, e.g., the price of the underlying asset or the volatility of the underlying asset, are important for hedging purposes. They change are called "the Greeks" of the portfolio, and are denoted by symbols from the Greek alphabet. The most widely used Greeks are listed below.

• Delta (Δ) is the rate of change of the value of the portfolio with respect to the spot price S of the underlying asset:

$$\Delta(V) \;=\; \frac{\partial V}{\partial S}; \tag{3.61}$$

• Gamma (Γ) is the rate of change of the Delta of the portfolio with respect to the spot price S of the underlying asset, i.e., the second partial derivative of the value of the portfolio with respect to S:

$$\Gamma(V) \;=\; \frac{\partial^2 V}{\partial S^2}; \tag{3.62}$$

• Theta (Θ) is the rate of change of the value of the portfolio with respect to time t (not with respect to maturity T):

$$\Theta(V) \;=\; \frac{\partial V}{\partial t}; \tag{3.63}$$

• Rho (ρ) is the rate of change of the value of the portfolio with respect to the risk–free interest rate r:

$$\rho(V) \;=\; \frac{\partial V}{\partial r}; \tag{3.64}$$

• Vega is the rate of change of the value of the portfolio with respect to the volatility σ of the underlying asset:

$$\mathrm{vega}(V) \;=\; \frac{\partial V}{\partial \sigma}. \tag{3.65}$$

From the Black–Scholes formulas (3.53) and (3.54) for options on assets paying continuous dividends, the following closed formulas for computing the Greeks of European plain vanilla call and put options can be derived:

[5]The portfolio could simply be, for example, the underlying asset, or a plain vanilla option on the underlying asset.

$$\Delta(C) = e^{-q(T-t)}N(d_1); \tag{3.66}$$
$$\Delta(P) = -e^{-q(T-t)}N(-d_1); \tag{3.67}$$

$$\Gamma(C) = \frac{e^{-q(T-t)}}{S\sigma\sqrt{T-t}}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}}; \tag{3.68}$$
$$\Gamma(P) = \Gamma(C); \tag{3.69}$$

$$\text{vega}(C) = Se^{-q(T-t)}\sqrt{T-t}\,\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}}; \tag{3.70}$$
$$\text{vega}(P) = \text{vega}(C); \tag{3.71}$$

$$\Theta(C) = -\frac{S\sigma e^{-q(T-t)}}{2\sqrt{2\pi(T-t)}}e^{-\frac{d_1^2}{2}} + qSe^{-q(T-t)}N(d_1) - rKe^{-r(T-t)}N(d_2); \tag{3.72}$$
$$\Theta(P) = -\frac{S\sigma e^{-q(T-t)}}{2\sqrt{2\pi(T-t)}}e^{-\frac{d_1^2}{2}} - qSe^{-q(T-t)}N(-d_1) + rKe^{-r(T-t)}N(-d_2); \tag{3.73}$$

$$\rho(C) = K(T-t)e^{-r(T-t)}N(d_2); \tag{3.74}$$
$$\rho(P) = -K(T-t)e^{-r(T-t)}N(-d_2). \tag{3.75}$$

The computations for deriving these formulas are based on differentiating improper integrals with respect to the integration limits; cf. section 2.3.

Note that, as soon as the formulas for the Greeks corresponding to a call option are derived, the Put–Call parity can be used to obtain the corresponding Greeks for the put option. For example, to compute the Delta of a put option, we write the Put–Call parity formula (1.47) as

$$P = C - Se^{-q(T-t)} + Ke^{-r(T-t)}. \tag{3.76}$$

By differentiating (3.76) with respect to S, we find that

$$\Delta(P) = \Delta(C) - e^{-q(T-t)}.$$

Using formula (3.66) for $\Delta(C)$ and Lemma 3.12, we obtain the formula (3.67) for $\Delta(P)$, i.e.,

$$\Delta(P) = e^{-q(T-t)}N(d_1) - e^{-q(T-t)} = -e^{-q(T-t)}(1-N(d_1)) = -e^{-q(T-t)}N(-d_1).$$

While definitions (3.61–3.65) are valid for the Greeks of any option on one underlying asset, including, e.g., American and exotic options, closed formulas for the Greeks can only be obtained if closed formulas for pricing the options exist. This happens very rarely for options that are not plain vanilla European, a notable exception being European barrier options; see section 7.8.

3.6.1 Explaining the magic of Greeks computations

It is interesting to note that the formulas (3.66–3.75) for the Greeks are simpler than expected. For example, the Delta of a call option is defined as

$$\Delta(C) = \frac{\partial C}{\partial S}.$$

Differentiating the Black–Scholes formula (3.53) with respect to S, we obtain

$$\Delta(C) = e^{-q(T-t)}N(d_1) + Se^{-q(T-t)}\frac{\partial}{\partial S}(N(d_1)) - Ke^{-r(T-t)}\frac{\partial}{\partial S}(N(d_2)), \tag{3.77}$$

since both d_1 and d_2 are functions of S; cf. (3.55) and (3.56).

However, we know from (3.66) that

$$\Delta(C) = e^{-q(T-t)}N(d_1). \tag{3.78}$$

To understand how (3.77) reduces to (3.78), we apply chain rule and obtain that

$$\frac{\partial}{\partial S}(N(d_1)) = N'(d_1)\frac{\partial d_1}{\partial S}; \tag{3.79}$$
$$\frac{\partial}{\partial S}(N(d_2)) = N'(d_2)\frac{\partial d_2}{\partial S}. \tag{3.80}$$

Then, using (3.79) and (3.80), we can write (3.77) as

$$\Delta(C) = e^{-q(T-t)}N(d_1) + Se^{-q(T-t)}N'(d_1)\frac{\partial d_1}{\partial S} - Ke^{-r(T-t)}N'(d_2)\frac{\partial d_2}{\partial S}. \tag{3.81}$$

Note that formulas (3.55) and (3.56) for d_1 and d_2 can be written as

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r-q)(T-t)}{\sigma\sqrt{T-t}} + \frac{\sigma\sqrt{T-t}}{2}; \tag{3.82}$$
$$d_2 = d_1 - \sigma\sqrt{T-t} = \frac{\ln\left(\frac{S}{K}\right) + (r-q)(T-t)}{\sigma\sqrt{T-t}} - \frac{\sigma\sqrt{T-t}}{2}. \tag{3.83}$$

The following result explains why (3.81) reduces to (3.78):

Lemma 3.15. *Let d_1 and d_2 be given by (3.82) and (3.83). Then*

$$Se^{-q(T-t)}\ N'(d_1) \;=\; Ke^{-r(T-t)}\ N'(d_2). \tag{3.84}$$

Proof. Recall that $N(z)$ is the cumulative distribution of the standard normal variable, i.e.,

$$N(z) \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}\,dx.$$

From Lemma 2.3, we find that $N'(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$. Then,

$$N'(d_1) \;=\; \frac{1}{\sqrt{2\pi}}\,e^{\frac{-d_1^2}{2}}; \tag{3.85}$$

$$N'(d_2) \;=\; \frac{1}{\sqrt{2\pi}}\,e^{\frac{-d_2^2}{2}}. \tag{3.86}$$

Therefore, in order to prove (3.84), it is enough to show that the following formula holds true:

$$Se^{-q(T-t)}\ e^{-\frac{d_1^2}{2}} \;=\; Ke^{-r(T-t)}\ e^{-\frac{d_2^2}{2}}. \tag{3.87}$$

Recall the notation $\exp(x) = e^x$. Since $\exp(\ln(x)) = x$, we find that

$$\begin{aligned} Se^{-q(T-t)} &= Ke^{-r(T-t)}\ \frac{S}{K}e^{(r-q)(T-t)} \\ &= Ke^{-r(T-t)}\ \exp\left(\ln\left(\frac{S}{K}\right) + (r-q)(T-t)\right). \end{aligned}$$

From (3.82), it is easy to see that

$$\ln\left(\frac{S}{K}\right) + (r-q)(T-t) \;=\; d_1\sigma\sqrt{T-t} \;-\; \frac{\sigma^2(T-t)}{2},$$

and therefore

$$Se^{-q(T-t)} \;=\; Ke^{-r(T-t)}\ \exp\left(d_1\sigma\sqrt{T-t} \;-\; \frac{\sigma^2(T-t)}{2}\right).$$

Using the fact that $d_2 = d_1 - \sigma\sqrt{T-t}$, we obtain (3.87) as follows:

$$\begin{aligned} Se^{-q(T-t)}\ e^{-\frac{d_1^2}{2}} &= Ke^{-r(T-t)}\ \exp\left(-\frac{d_1^2}{2} \;+\; d_1\sigma\sqrt{T-t} \;-\; \frac{\sigma^2(T-t)}{2}\right) \\ &= Ke^{-r(T-t)}\ \exp\left(-\frac{(d_1-\sigma\sqrt{T-t})^2}{2}\right) \\ &= Ke^{-r(T-t)}\ e^{-\frac{d_2^2}{2}}. \end{aligned}$$

□

We return our attention to proving formula (3.78) for $\Delta(C)$.
From (3.82) and (3.83), we find that

$$\frac{\partial d_1}{\partial S} \;=\; \frac{\partial d_2}{\partial S} \;=\; \frac{1}{\sigma S\sqrt{T-t}}. \tag{3.88}$$

Using (3.88) and Lemma 3.15, we conclude that formula (3.81) becomes

$$\begin{aligned} \Delta(C) &= e^{-q(T-t)}N(d_1) \;+\; Se^{-q(T-t)}N'(d_1)\frac{\partial d_1}{\partial S} \;-\; Ke^{-r(T-t)}N'(d_2)\frac{\partial d_2}{\partial S} \\ &= e^{-q(T-t)}N(d_1) \;+\; Se^{-q(T-t)}\ N'(d_1)\left(\frac{\partial d_1}{\partial S} \;-\; \frac{\partial d_2}{\partial S}\right) \\ &= e^{-q(T-t)}N(d_1). \end{aligned}$$

Formula (3.78) is therefore proven.

The simplified formulas (3.70), (3.72), and (3.74) for the vega, Θ, and ρ of a European call option[6] are obtained similarly using Lemma 3.15.

The formula for vega(C):
We differentiate the Black–Scholes formula (3.53) with respect to σ. Following the same steps as in the computation for the Delta of the call option, i.e., using chain rule and Lemma 2.3, we obtain that

$$\text{vega}(C) \;=\; \frac{\partial C}{\partial \sigma} \;=\; Se^{-q(T-t)}\ N'(d_1)\ \frac{\partial d_1}{\partial \sigma} \;-\; Ke^{-r(T-t)}\ N'(d_2)\ \frac{\partial d_2}{\partial \sigma}.$$

Using the result of Lemma 3.15, we conclude that

$$\text{vega}(C) \;=\; Se^{-q(T-t)}\ N'(d_1)\ \left(\frac{\partial d_1}{\partial \sigma} \;-\; \frac{\partial d_2}{\partial \sigma}\right).$$

Since $d_2 = d_1 - \sigma\sqrt{T-t}$, we find that $d_1 - d_2 = \sigma\sqrt{T-t}$ and thus

$$\frac{\partial d_1}{\partial \sigma} \;-\; \frac{\partial d_2}{\partial \sigma} \;=\; \sqrt{T-t}.$$

Then, using the fact that $N'(d_1) = \frac{1}{\sqrt{2\pi}}\,e^{\frac{-d_1^2}{2}}$, see (3.85), we conclude that

$$\text{vega}(C) \;=\; \frac{1}{\sqrt{2\pi}}\ Se^{-q(T-t)}\ e^{-\frac{d_1^2}{2}}\ \sqrt{T-t},$$

[6]Note that the formulas (3.71), (3.73), and (3.75) for the vega, Θ, and ρ of a European put option can be obtained from (3.70), (3.72), and (3.74) by using the Put–Call parity.

which is the same as formula (3.70).

The formula for $\Theta(C)$:
We differentiate the Black–Scholes formula (3.53) with respect to t. Using chain rule and Lemma 2.3 we obtain that

$$\begin{aligned}\Theta(C) &= Se^{-q(T-t)}\, N'(d_1)\, \frac{\partial d_1}{\partial t} + qSe^{-q(T-t)}N(d_1)\\ &\quad - Ke^{-r(T-t)}\, N'(d_2)\, \frac{\partial d_2}{\partial t} - rKe^{-r(T-t)}N(d_2).\end{aligned}$$

Using the result of Lemma 3.15, we conclude that

$$\begin{aligned}\Theta(C) &= Se^{-q(T-t)}\, N'(d_1) \left(\frac{\partial d_1}{\partial t} - \frac{\partial d_2}{\partial t}\right)\\ &\quad + qSe^{-q(T-t)}N(d_1) - rKe^{-r(T-t)}N(d_2).\end{aligned}$$

Since $d_1 - d_2 = \sigma\sqrt{T-t}$, we find that

$$\frac{\partial d_1}{\partial t} - \frac{\partial d_2}{\partial t} = -\frac{\sigma}{2\sqrt{T-t}}.$$

Since $N'(d_1) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-d_1^2}{2}}$, cf. (3.85), we conclude that

$$\begin{aligned}\Theta(C) &= -\frac{1}{\sqrt{2\pi}}\, Se^{-q(T-t)}\, e^{-\frac{d_1^2}{2}}\, \frac{\sigma}{2\sqrt{T-t}}\\ &\quad + qSe^{-q(T-t)}N(d_1) - rKe^{-r(T-t)}N(d_2),\end{aligned}$$

which is the same as formula (3.72).

The formula for $\rho(C)$:
We differentiate the Black–Scholes formula (3.53) with respect to r. Using Chain Rule and Lemma 2.3 we obtain that

$$\rho(C) = Se^{-q(T-t)}N'(d_1)\frac{\partial d_1}{\partial r} - Ke^{-r(T-t)}N'(d_2)\frac{\partial d_2}{\partial r} + K(T-t)e^{-r(T-t)}N(d_2).$$

Using Lemma 3.15, we find that

$$\rho(C) = Se^{-q(T-t)}\, N'(d_1) \left(\frac{\partial d_1}{\partial r} - \frac{\partial d_2}{\partial r}\right) + K(T-t)e^{-r(T-t)}N(d_2).$$

Since $d_1 - d_2 = \sigma\sqrt{T-t}$, we find that

$$\frac{\partial d_1}{\partial r} = \frac{\partial d_2}{\partial r},$$

and therefore,

$$\rho(C) = K(T-t)e^{-r(T-t)}N(d_2),$$

which is the same as formula (3.74).

3.6.2 Implied volatility

The only parameter needed in the Black–Scholes formulas (3.53–3.56) that is not directly observable in the markets is the volatility σ of the underlying asset. The risk free rate r and the continuous dividend yield q of the asset can be estimated from market data; the maturity date T and the strike K of the option, as well as the spot price S of the underlying asset are known when a price for the option is quoted.

The implied volatility σ_{imp} is the value of the volatility parameter σ that makes the Black–Scholes value of the option equal to the traded price of the option. To formalize this, denote by $C_{BS}(S,K,T,\sigma,r,q)$ the Black–Scholes value of a call option with strike K and maturity T on an underlying asset with spot price S paying dividends continuously at the rate q, if interest rates are constant and equal to r. If C is the quoted values of a call with parameters S, K, T, σ, r, and q, the implied volatility σ_{imp} corresponding to the price C is, by definition, the solution to

$$C_{BS}(S,K,T,\sigma_{imp},r,q) = C. \tag{3.89}$$

The implied volatility can also be derived from the given price P of a put option, by solving

$$P_{BS}(S,K,T,\sigma_{imp},r,q) = P, \tag{3.90}$$

where $P_{BS}(S,K,T,\sigma,r,q)$ is the Black–Scholes value of a put option.

Note that, as functions of volatility, the Black–Scholes values of both call and put options are strictly increasing since

$$\begin{aligned}\frac{\partial C_{BS}}{\partial \sigma} = \frac{\partial P_{BS}}{\partial \sigma} &= \mathrm{vega}(C_{BS}) = \mathrm{vega}(P_{BS})\\ &= Se^{-qT}\sqrt{T}\, \frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}} > 0;\end{aligned} \tag{3.91}$$

cf. (3.70) and (3.71). (Throughout this section, to keep notation simple, we assume that the present time is $t = 0$.)

Therefore, if a solution σ_{imp} for (3.89) exists, it will be unique, and the implied volatility will be well defined. Similarly, equation (3.90) has at most one solution.

For the implied volatility to exist and be nonnegative, the given value C of the call option must be arbitrage–free, i.e.,

$$Se^{-qT} - Ke^{-rT} \le C < Se^{-qT}. \tag{3.92}$$

The bounds for the call option price from (3.92) can be obtained by using the Law of One Price; cf. Theorem 1.10.

Similarly, the implied volatility derived from (3.90) exists and has nonnegative value if and only if the value P of the put satisfies the following inequality

$$Ke^{-rT} - Se^{-qT} \leq P < Ke^{-rT}.$$

From a computational perspective, problems (3.89) and (3.90) are one dimensional nonlinear equations and can be solved very efficiently using Newton's method; see section 8.2.2 for more details.

We conclude this section by showing that the implied volatilities corresponding to put and call options with the same strike and maturity on the same underlying asset must be equal.

Denote by $\sigma_{imp,P}$ and $\sigma_{imp,C}$ the implied volatilities corresponding to a put option with price P and to a call option with price C, respectively. Both options have strike K and maturity T and are written on the same underlying asset. In other words,

$$P_{BS}(\sigma_{imp,C}) = P \quad \text{and} \quad C_{BS}(\sigma_{imp,C}) = C, \tag{3.93}$$

where we denoted $P_{BS}(S, K, T, \sigma_{imp,C}, r, q)$ and $C_{BS}(S, K, T, \sigma_{imp,C}, r, q)$ by $P_{BS}(\sigma_{imp,C})$ and $C_{BS}(\sigma_{imp,C})$, respectively. For no–arbitrage, the option values P and C must satisfy the Put–Call parity (1.47), i.e.,

$$P + Se^{-qT} - C = Ke^{-rT}. \tag{3.94}$$

From (3.93) and (3.94) we find that

$$P_{BS}(\sigma_{imp,P}) + Se^{-qT} - C_{BS}(\sigma_{imp,C}) = Ke^{-rT}. \tag{3.95}$$

Recall that the Black–Scholes values of put and call options satisfy the Put–Call parity, i.e.,

$$P_{BS}(\sigma) + Se^{-qT} - C_{BS}(\sigma) = Ke^{-rT}, \tag{3.96}$$

for any value $\sigma > 0$ of the volatility. Let $\sigma = \sigma_{imp,C}$ in (3.96). Then,

$$P_{BS}(\sigma_{imp,C}) + Se^{-qT} - C_{BS}(\sigma_{imp,C}) = Ke^{-rT}. \tag{3.97}$$

From (3.95) and (3.97) it follows that

$$P_{BS}(\sigma_{imp,P}) = P_{BS}(\sigma_{imp,C}).$$

As mentioned before, the Black–Scholes value of a put option is a strictly increasing function of volatility; cf. (3.91). Therefore, we obtain that

$$\sigma_{imp,P} = \sigma_{imp,C},$$

and conclude that the implied volatilities corresponding to put and call options with the same strike and maturity on the same asset must be equal.

3.7 The concept of hedging. Δ–hedging and Γ–hedging

Assume you are long a call option. If the price of the underlying asset declines, the value of the call decreases and the long call position loses money. To protect against a downturn in the price of the underlying asset (i.e., to hedge), you sell short[7] Δ units of the underlying asset; note that Δ is a number in this context. You now have a portfolio consisting of a long position in one call option and a short position in Δ units of the underlying asset. The goal is to choose Δ in such a way that the value of the portfolio is not sensitive to small changes in the price of the underlying asset.

If Π is the value of the portfolio, then $\Pi = C - \Delta \cdot S$, or, equivalently,

$$\Pi(S) = C(S) - \Delta \cdot S.$$

Assume that the spot price of the underlying asset changes to $S + dS$, where dS is small, i.e., $dS << S$. The change in the value of the portfolio is

$$\begin{aligned} \Pi(S + dS) - \Pi(S) &= C(S + dS) - \Delta \cdot (S + dS) - (C(S) - \Delta \cdot S) \\ &= C(S + dS) - C(S) - \Delta\, dS. \end{aligned} \tag{3.98}$$

We look for Δ such that the value of the portfolio is insensitive to small changes in the price of the underlying asset, i.e., such that

$$\Pi(S + dS) - \Pi(S) \approx 0. \tag{3.99}$$

From (3.98) and (3.99), and solving for Δ, we find that

$$\Delta \approx \frac{C(S + dS) - C(S)}{dS}.$$

[7]To explain short selling, consider the case of equity options, i.e., options where the underlying asset is stock. Selling short one share of stock is done by borrowing the share (through a broker), and then selling the share on the market. Part of the cash is deposited with the broker in a margin account as collateral (usually, 50% of the sale price), while the rest is deposited in a brokerage account. The margin account must be settled when a margin call is issued, which happens when the price of the shorted asset appreciates beyond a certain level. Cash must then be added to the margin account to reach the level of 50% of the amount needed to close the short, i.e., to buy one share of stock at the current price of the asset, or the short is closed by the broker on your behalf. The cash from the brokerage account can be invested freely, while the cash from the margin account earns interest at a fixed rate, but cannot be invested otherwise. The short is closed by buying the share (at a later time) on the market and returning it to the original owner (via the broker; the owner rarely knows that the asset was borrowed and sold short).

We will not consider here these or other issues, such as margin calls, the liquidity of the market and the availability of shares for short selling, transaction costs, and the impossibility of taking the exact position required for the "correct" hedge.

By letting $dS \to 0$, we find that the appropriate position Δ in the underlying asset in order to hedge a call option is

$$\Delta = \frac{\partial C}{\partial S},$$

which is the same as $\Delta(C)$, the Delta of a call option defined in (3.61) as the rate of change of the value of the call with respect to changes in the price of the underlying asset.

A similar argument on hedging a long position in a portfolio with price V made of derivative securities on a single underlying asset shows that the appropriate Δ–hedging position to be taken in the asset is

$$\Delta = \frac{\partial V}{\partial S},$$

where S is the spot price of the asset. The value of the new portfolio,

$$\Pi = V - \Delta \cdot S,$$

will be insensitive to small changes in the price of the underlying asset. Such a portfolio is called Delta–neutral, since

$$\Delta(\Pi) = \frac{\partial \Pi}{\partial S} = \frac{\partial V}{\partial S} - \Delta = 0.$$

Recall that the correct hedging position for a long call is short Δ shares. Therefore, the hedging position for a short call is long Δ shares. Similarly, a long put position is hedged by going long Δ shares: as the spot price of the underlying asset goes up, a long put position loses value while a long asset position will gain value, thus offsetting the loss on the long put. A short put is hedged by going short Δ shares.

Option Position	Hedge Position	Sign of Option Delta
Long Call	Short Asset	Positive Delta
Short Call	Long Asset	Negative Delta
Long Put	Long Asset	Negative Delta
Short Put	Short Asset	Positive Delta

Note that a portfolio is Delta–neutral only over a short period of time. As the price of the underlying asset changes, the portfolio might become unbalanced, i.e., not Delta–neutral. In this case, the hedge needs to be rebalanced, i.e., units of the underlying asset must be bought or sold in order to make the portfolio Delta–neutral again. Deciding when to rebalance the portfolio is a compromise between rebalancing the hedge often (which

keeps the portfolio close to Delta–neutral, and therefore reduces the risk of losses if the price of the underlying asset changes rapidly, but increases the trading costs), and rebalancing only when the hedge becomes inaccurate enough (which reduces the trading costs but increases the risk of having to manage a portfolio that is close to, but not exactly Delta–neutral, most of the time).

To achieve a better hedge of the portfolio, i.e., to have a portfolio that is even less sensitive to small changes in the price of the underlying asset than a Delta–neutral portfolio, we look for a portfolio that is both Delta–neutral and Gamma–neutral, i.e., such that

$$\Delta(\Pi) = \frac{\partial \Pi}{\partial S} = 0 \quad \text{and} \quad \Gamma(\Pi) = \frac{\partial^2 \Pi}{\partial S^2} = 0,$$

where the Gamma of the portfolio is denoted by $\Gamma(\Pi)$. Note that

$$\Gamma(\Pi) = \frac{\partial \Delta(\Pi)}{\partial S}.$$

Therefore, if a portfolio is Gamma–neutral, its Delta will be rather insensitive to small changes in the price of the underlying asset. In particular, if the portfolio is both Delta–neutral and Gamma–neutral, then the Delta of the portfolio will change significantly only if larger changes in the price of the underlying asset occur. The need to rebalance the hedge for such a portfolio occurs less often than in the case of a portfolio that is Delta–neutral but not Gamma–neutral. This saves trading costs and the portfolios are, generally speaking, better hedged most of the time.

To obtain a Delta– and Gamma–neutral portfolio, starting with a portfolio V made of derivative securities over a single underlying asset, we need to be able to trade in two different derivative securities on the underlying asset, e.g., the asset itself and an option (call or put, with any maturity) on the asset. Let $\Delta(V)$ and $\Gamma(V)$ be the Delta and Gamma of the portfolio, respectively. Assume that we can take long and short positions of arbitrary size in two derivative securities with values D_1 and D_2 on the same underlying asset. The value Π of a portfolio made of the initial portfolio and positions x_1 and x_2 in D_1 and D_2, respectively, is

$$\Pi = V + x_1 D_1 + x_2 D_2.$$

This portfolio is Delta– and Gamma–neutral, i.e., $\Delta(\Pi) = 0$ and $\Gamma(\Pi) = 0$, if and only if

$$\begin{aligned} \Delta(\Pi) &= \Delta(V) + x_1 \Delta(D_1) + x_2 \Delta(D_2) = 0; \\ \Gamma(\Pi) &= \Gamma(V) + x_1 \Gamma(D_1) + x_2 \Gamma(D_2) = 0. \end{aligned}$$

Then, x_1 and x_2 are the solutions of the linear system

$$\begin{cases} x_1\Delta(D_1) + x_2\Delta(D_2) = -\Delta(V); \\ x_1\Gamma(D_1) + x_2\Gamma(D_2) = -\Gamma(V). \end{cases}$$

Note that this system has a unique solution as long as D_1 and D_2 do not have the same Delta to Gamma ratio, i.e., if

$$\frac{\Delta(D_1)}{\Gamma(D_1)} \neq \frac{\Delta(D_2)}{\Gamma(D_2)}.$$

Otherwise, a position in another derivative security with a different Delta–to–Gamma ratio will have to be taken in order to make the portfolio both Delta–neutral and Gamma–neutral.

Example: Assume you hold a portfolio Π with Delta equal to 1500, i.e., $\Delta(\Pi) = 1500$. The portfolio contains derivative securities depending on only one asset. You can take positions in, e.g., a call option on the same asset with Delta equal to 0.3. What position should you take in the call option to make the portfolio Delta–neutral?

Answer: Let x be the size of your position in the call option, and let C be the price of the call. (Depending on the sign of x, you are long x calls if $x > 0$, and short x calls if $x < 0$.) The value Π_{new} of the new portfolio is $\Pi_{new} = \Pi + xC$, while the Delta of Π_{new} is

$$\Delta(\Pi_{new}) = \Delta(\Pi) + x\Delta(C) = 1500 + 0.3\,x.$$

Thus, $\Delta(\Pi_{new}) = 0$ if and only if $x = -5000$.

You must short 5000 calls (or sell 50 call options contracts, if one contract is written for 100 options) to make the portfolio Delta–neutral. □

3.8 Implementation of the Black–Scholes formula

To implement the Black–Scholes formulas (3.53–3.54), the cumulative distribution $N(t)$ of the standard normal variable Z must be estimated numerically. The most widely used algorithm for approximating $N(t)$ has an approximation error that is smaller than $7.5 \cdot 10^{-7}$ at any point on the real axis; cf. Abramowitz and Stegun [1], page 932. We call this routine cum_dist_normal(tt), and include its pseudocode in Table 3.1.

The pseudocode for implementing the Black–Scholes formulas using the routine cum_dist_normal(t) for approximating $N(t)$, is given in Table 3.2.

Table 3.1: Pseudocode for computing the cumulative distribution of Z

```
Input:
t = real number

Output:
function nn = cum_dist_normal(t)

z = abs(t)
y = 1/(1 + 0.2316419z)
a1 = 0.319381530; a2 = −0.356563782; a3 = 1.781477937
a4 = −1.821255978; a5 = 1.330274429
m = 1 − exp(−t^2/2)(a1 y + a2 y^2 + a3 y^3 + a4 y^4 + a5 y^5)/sqrt(2π)
if t > 0
   nn = m
else
   nn = 1 − m
end
```

Table 3.2: Pseudocode for Black–Scholes formula

```
Input:
t = present time (often equal to 0)
S = spot price of the underlying asset (at time t)
K = option strike
T = maturity date (time to maturity is T − t)
σ = volatility of the underlying asset
r = constant interest rate
q = continuous dividend rate of the underlying asset

Output:
C = price of the European call option
P = price of the European put option
```

$$d_1 = \left(\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)(T-t)\right) / \left(\sigma\sqrt{T-t}\right);\ d_2 = d_1 - \sigma\sqrt{T-t}$$
$$C = Se^{-q(T-t)}\text{cum\_dist\_normal}(d_1) - Ke^{-r(T-t)}\text{cum\_dist\_normal}(d_2)$$
$$P = Ke^{-r(T-t)}\text{cum\_dist\_normal}(-d_2) - Se^{-q(T-t)}\text{cum\_dist\_normal}(-d_1)$$

Example: Use the Black–Scholes formula to price a six months European call option with strike 40, on an underlying asset with spot price 42 and volatility 30%, which pays dividends continuously, with dividend rate 3%. Assume that interest rates are constant at 5%.

Price a six months European put option with strike 40 on the same asset, using the Black–Scholes formula. Check whether the Put–Call parity is satisfied.

Answer: Let $t = 0$ be the present time. The maturity of the option is $T = 0.5$, computed in years. We know the spot price of the underlying asset, the strike price and maturity of the options, the volatility and dividend yield of the underlying asset, and the constant interest rate. This is exactly the input required by the Black–Scholes implementation from Table 3.2, i.e.,

$$t = 0;\ S = 42;\ K = 40;\ T = 0.5;\ \sigma = 0.3;\ q = 0.03;\ r = 0.05.$$

We obtain that $d_1 = 0.383206$ and $d_2 = 0.171073$, and

$$\mathrm{cum\_dist\_normal}(d_1) = 0.649216; \quad \mathrm{cum\_dist\_normal}(d_2) = 0.567917.$$

The Black–Scholes price of the call is $C = 4.705325$ and the price of the put is $P = 2.343022$. Note that

$$P + Se^{-qT} - C - Ke^{-rT} = 2.1738 \cdot 10^{-6}.$$

Since we used the approximate routine cum_dist_normal to compute the cumulative distribution of the standard normal variable, an error of this magnitude in the Put–Call parity (1.47) was to be expected. □

3.9 References

A good introductory probability book is Ross [21]; for a measure based introduction to probability, see Capinski and Kopp [4], and also Durrett [9].

The Black–Scholes formula and the lognormal model for the evolution of the underlying asset are presented in great detail and with pedagogical insights in Chriss [5]. More Black–Scholes type formulas, as well as many other options pricing formulas can be found in Haug [12].

The numerical implementations of various pricing methods for options, including binomial trees, are presented in Clewlow and Strickland [6]. Taleb [31] gives a very practical perspective on dynamic hedging.

3.10 Exercises

1. Let k be a positive integer with $2 \leq k \leq 12$. You throw two fair dice. If the sum of the dice is k, you win $w(k)$, or lose 1 otherwise. Find the smallest value of $w(k)$ thats makes the game worth playing.

2. A coin lands heads with probability p and tails with probability $1 - p$. Let X be the number of times you must flip the coin before it lands heads. What are $E[X]$ and $\mathrm{var}(X)$?

3. Recall the stock evolution model from section 3.1: over each of three consecutive time intervals of length $\tau = 1/12$, a stock with spot price $S_0 = 40$ at time $t = 0$ will either go up by a factor $u = 1.05$ with probability $p = 0.6$, or down by a factor $d = 0.96$ with probability $1 - p = 0.4$. Compute the expected value and the variance of the stock price at time $T = 3\tau$, i.e., compute $E[S_T]$ and $\mathrm{var}(S_T)$.

4. The density function of the exponential random variable X with parameter $\alpha > 0$ is

$$f(x) = \begin{cases} \alpha\, e^{-\alpha x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

(i) Show that the function $f(x)$ is indeed a density function. It is clear that $f(x) \geq 0$, for any $x \in \mathbb{R}$. Prove that

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

(ii) Show that the expected value and the variance of the exponential random variable X are $E[X] = \frac{1}{\alpha}$ and $\mathrm{var}(X) = \frac{1}{\alpha^2}$.

(iii) Show that the cumulative density of X is

$$F(x) = \begin{cases} 1 - e^{-\alpha x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

(iv) Show that

$$P(X \geq t) = \int_t^{\infty} f(x)\, dx = e^{\alpha t}.$$

Note: this result is used to show that the exponential variable is memoryless, i.e., $P(X \geq t + s \mid X \geq t) = P(X \geq s)$.

5. Use the technique from the proof of Lemma 3.8 to show that

$$\int_a^b f(x)g(x)\,dx \;\le\; \left(\int_a^b f^2(x)\,dx\right)^{\frac{1}{2}} \left(\int_a^b g^2(x)\,dx\right)^{\frac{1}{2}},$$

for any two continuous functions $f, g : \mathbb{R} \to \mathbb{R}$.

Hint: Use the fact that

$$\int_a^b (f(x) + \alpha g(x))^2\,dx \;\ge\; 0, \;\; \forall\, \alpha \in \mathbb{R}.$$

6. Use the Black–Scholes formula to price both a put and a call option with strike $K = 45$ expiring in six months on an underlying asset with spot price 50 and volatility 20% paying dividends continuously at 2%, if interest rates are constant at 6%.

7. What is the value of a European Put option with strike $K = 0$? What is the value of a European Call option with strike $K = 0$? How do you hedge a short position in such a call option?

Note: Since no assumptions are made on the evolution of the price of the underlying asset, the Law of One Price from section 1.8, and not the Black–Scholes formulas, must be used.

8. Use formula (3.74), i.e., $\rho(C) = K(T-t)e^{-r(T-t)}N(d_2)$, and the Put–Call parity to obtain the formula (3.75) for $\rho(P)$, i.e.,

$$\rho(P) \;=\; -\,K(T-t)e^{-r(T-t)}N(-d_2).$$

9. The sensitivity of the vega of a portfolio with respect to volatility and to the price of the underlying asset are often important to estimate, e.g., for pricing volatility swaps. These two Greeks are called volga and vanna and are defined as follows:

$$\text{volga}(V) \;=\; \frac{\partial(\text{vega}(V))}{\partial\sigma} \quad \text{and} \quad \text{vanna}(V) \;=\; \frac{\partial(\text{vega}(V))}{\partial S}.$$

It is easy to see that

$$\text{volga}(V) \;=\; \frac{\partial^2 V}{\partial\sigma^2} \quad \text{and} \quad \text{vanna}(V) \;=\; \frac{\partial^2 V}{\partial S\partial\sigma}.$$

The name volga is the short for "volatility gamma". Also, vanna can be interpreted as the rate of change of the Delta with respect to the volatility of the underlying asset, i.e.,

$$\text{vanna}(V) \;=\; \frac{\partial(\Delta(V))}{\partial\sigma}.$$

(i) Compute the volga and vanna for a plain vanilla European call option on an asset paying dividends continuously at the rate q.

(ii) Use the Put–Call parity to compute the volga and vanna for a plain vanilla European put option.

10. Show that an ATM call on an underlying asset paying dividends continuously at rate q is worth more than an ATM put with the same maturity if and only if $q \le r$, where r is the constant risk free rate. Use the Put–Call parity, and then use the Black–Scholes formula to prove this result.

11. (i) Show that the Theta of a plain vanilla European call option on a non–dividend–paying asset is always negative.

(ii) For long dated (i.e., with $T-t$ large) ATM calls on an underlying asset paying dividends continuously at a rate equal to the constant risk–free rate, i.e., with $q = r$, show that the Theta may be positive.

12. Show that the price of a plain vanilla European call option is a convex function of the strike of the option, i.e., show that

$$\frac{\partial^2 C}{\partial K^2} \;\ge\; 0.$$

Hint: Use the result of Lemma 3.15 to obtain that

$$\frac{\partial C}{\partial K} \;=\; -\,e^{-r(T-t)}N(d_2).$$

13. Compute the Gamma of ATM call options with maturities of fifteen days, three months, and one year, respectively, on a non–dividend–paying underlying asset with spot price 50 and volatility 30%. Assume that interest rates are constant at 5%. What can you infer about the hedging of ATM options with different maturities?

14. (i) It is easy to see from (3.70) and (3.71), i.e., from

$$\text{vega}(C) = \text{vega}(P) = Se^{-q(T-t)}\sqrt{T-t}\,\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}},$$

that the vega of a plain vanilla European call or put is positive. Can you give a financial explanation for this?

(ii) Compute the vega of ATM Call options with maturities of fifteen days, three months, and one year, respectively, on a non–dividend–paying underlying asset with spot price 50 and volatility 30%. For simplicity, assume zero interest rates, i.e., $r = 0$.

(iii) If $r = q = 0$, the vega of ATM call and put options is

$$\text{vega}(C) = \text{vega}(P) = S\sqrt{T-t}\,\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}},$$

where $d_1 = \frac{\sigma\sqrt{T-t}}{2}$. Compute the dependence of vega(C) on time to maturity $T - t$, i.e.,

$$\frac{\partial\,(\text{vega}(C))}{\partial(T-t)},$$

and explain the results from part (ii) of the problem.

15. Assume that interest rates are constant and equal to r. Show that, unless the price of a call option C with strike K and maturity T on a non–dividend paying asset with spot price S satisfies the inequality

$$Se^{-qT} - Ke^{-rT} \leq C \leq Se^{-qT},$$

arbitrage opportunities arise.

Show that the value P of the corresponding put option must satisfy the following no–arbitrage condition:

$$Ke^{-rT} - Se^{-qT} \leq P \leq Ke^{-rT}.$$

16. A portfolio containing derivative securities on only one asset has Delta 5000 and Gamma -200. A call on the asset with $\Delta(C) = 0.4$ and $\Gamma(C) = 0.05$, and a put on the same asset, with $\Delta(P) = -0.5$ and $\Gamma(P) = 0.07$ are currently traded. How do you make the portfolio Delta–neutral and Gamma–neutral?

17. You are long 1000 call options with strike 90 and three months to maturity. Assume that the underlying asset has a lognormal distribution with drift $\mu = 0.08$ and volatility $\sigma = 0.2$, and that the spot price of the asset is 92. The risk-free rate is $r = 0.05$. What Delta–hedging position do you need to take? (Note that, for Delta–hedging purposes, it is not necessary to know the drift μ of the underlying asset, since $\Delta(C)$ does not depend on μ; cf. (3.66).)

18. You buy 1000 six months ATM Call options on a non–dividend–paying asset with spot price 100, following a lognormal process with volatility 30%. Assume the interest rates are constant at 5%.

(i) How much money do you pay for the options?

(ii) What Delta–hedging position do you have to take?

(iii) On the next trading day, the asset opens at 102. What is the value of your position (the option and shares position)?

(iv) Had you not Delta–hedged, how much would you have lost due to the increase in the price of the asset?

Chapter 4

Lognormal random variables. Risk–neutral pricing.

Change of probability density function for functions of random variables.

Lognormal random variables.

Independent random variables.

Approximating sums of lognormal variables.

Convergence of power series. Radius of convergence. Stirling's formula.

4.1 Change of density function for functions of random variables

Let X be a random variable and let $f(x)$ be the density function of X. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $\int_{-\infty}^{\infty} |g(x)| dx < \infty$. Then $Y = g(X)$ is a random variable. Under certain assumptions, it is possible to compute the density function of Y given the density function of X.

Lemma 4.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be an invertible function which is differentiable and such that $\int_{-\infty}^{\infty} |g(x)| dx < \infty$. Let X be a random variable with density function $f(x)$ and let $Y = g(X)$. Then, the density function of Y is either*

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}, \tag{4.1}$$

if $g(x)$ is an increasing function, or

$$h(y) = -\frac{f(g^{-1}(y))}{g'(g^{-1}(y))}, \tag{4.2}$$

if $g(x)$ is a decreasing function.

Proof. Let $h(y)$ be the density function of Y. Then, by definition,

$$P(a \le Y \le b) = \int_a^b h(y)\,dy. \tag{4.3}$$

It is easy to see that any continuous function $g(x)$ that is invertible must be either strictly increasing, or strictly decreasing.

Assume that $g(x)$ is increasing. Since $Y = g(X)$, we find that

$$\begin{aligned} P(a \le Y \le b) &= P(a \le g(X) \le b) = P(g^{-1}(a) \le X \le g^{-1}(b)) \\ &= \int_{g^{-1}(a)}^{g^{-1}(b)} f(x)\,dx. \end{aligned} \tag{4.4}$$

We make the substitution $x = g^{-1}(y)$ in (4.4). The integral limits change from $x = g^{-1}(a)$ to $y = a$ and from $x = g^{-1}(b)$ to $y = b$. Also, from Lemma 1.1, it follows that

$$dx = \frac{1}{g'(g^{-1}(y))}\,dy.$$

Therefore, we find from (4.4) that

$$P(a \le Y \le b) = \int_a^b \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,dy. \tag{4.5}$$

The desired result (4.1) follows immediately from (4.3) and (4.5).

If $g(x)$ is decreasing, then (4.4) changes slightly, since in this case

$$P(a \le g(X) \le b) = P(g^{-1}(b) \le X \le g^{-1}(a)).$$

Thus, formula (4.4) becomes

$$P(a \le Y \le b) = \int_{g^{-1}(b)}^{g^{-1}(a)} f(x)\,dx.$$

After the same change of variables $x = g^{-1}(y)$, we find that

$$P(a \le Y \le b) = \int_b^a \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,dy = -\int_a^b \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,dy, \tag{4.6}$$

and (4.2) follows from (4.3) and (4.6). □

Note that it is usually easier to compute the expected value and variance of the random variable $Y = g(X)$ by regarding Y as a function of X and using Lemma 3.4, i.e.,

$$\begin{aligned} E[Y] &= E[g(X)] = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx; \\ \mathrm{var}(Y) &= E[Y^2] - (E[Y])^2 = \int_{-\infty}^{\infty} g^2(x)f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x)f(x)\,dx\right)^2, \end{aligned}$$

rather than using Lemma 4.1 to compute explicitely the probability density function of Y and using the definitions (3.9) and (3.13). An example of this approach is given in Lemma 4.3.

4.2 Lognormal random variables

To derive the Black–Scholes formulas, the underlying asset is assumed to follow a lognormal distribution, i.e.,

$$\ln\left(\frac{S(t_2)}{S(t_1)}\right) = \left(\mu - q - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma\sqrt{t_2 - t_1}Z, \ \ \forall\ 0 \le t_1 < t_2 \le T;$$

cf. (3.52). A random variable whose natural logarithm is a normal variable is therefore important for practical applications.

Definition 4.1. *A random variable Y is called a lognormal variable if there exists a normal variable X such that* $\ln(Y) = X$. *If $X = \mu + \sigma Z$, then*

$$\ln(Y) = \mu + \sigma Z, \tag{4.7}$$

and we say that Y is a lognormal variable with parameters μ and σ.

Lemma 4.2. *Let Y be a lognormal random variable with parameters μ and σ. The probability density function of Y is*

$$h(y) = \begin{cases} \frac{1}{y|\sigma|\sqrt{2\pi}}\, exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right), & \text{if } y > 0; \\ 0, & \text{if } y \le 0. \end{cases} \tag{4.8}$$

Proof. Let $X = \mu + \sigma Z$ such that $\ln(Y) = X$, where Z is the standard normal variable. Then $Y = e^X$. Therefore, $P(Y \le 0) = 0$. If $h(y)$ is the density function of Y, we obtain that $h(y) = 0$, for all $y \le 0$.

Since $Y = e^X$, we can write $Y = g(X)$, where $g(x) = e^x$ is an increasing invertible differentiable function on $\mathbb{R}$. Let $y > 0$. Note that

$$g^{-1}(y) = \ln(y) \quad \text{and} \quad g'(g^{-1}(y)) = g'(\ln(y)) = e^{\ln(y)} = y. \tag{4.9}$$

From Lemma 3.14, we know that the density function of $X = \mu + \sigma Z$ is

$$f(x) = \frac{1}{|\sigma|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{4.10}$$

From Lemma 4.1, and using (4.9) and (4.10), we find that, for any $y > 0$, the density function of Y is

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f(\ln(y))}{y} = \frac{1}{y|\sigma|\sqrt{2\pi}} \exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right).$$

□

Lemma 4.3. *The expected value and variance of a lognormal random variable Y with parameters μ and σ are*

$$E[Y] = e^{\mu+\frac{\sigma^2}{2}} \tag{4.11}$$

$$var(Y) = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right). \tag{4.12}$$

Proof. Note that $\ln(Y) = \mu + \sigma Z$. Then, $Y = e^{\mu+\sigma Z}$ and, using Lemma 3.4, we find that

$$E[Y] = E[e^{\mu+\sigma Z}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu+\sigma x}\, e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu+\sigma x-\frac{x^2}{2}}\, dx.$$

By completing the square on the exponent $\mu + \sigma x - \frac{x^2}{2}$, i.e., by writing

$$\mu + \sigma x - \frac{x^2}{2} = \mu + \frac{\sigma^2}{2} - \frac{(x-\sigma)^2}{2},$$

we obtain that

$$E[Y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu+\frac{\sigma^2}{2}-\frac{(x-\sigma)^2}{2}}\, dx = e^{\mu+\frac{\sigma^2}{2}} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\sigma)^2}{2}}\, dx\right). \tag{4.13}$$

Using the substitution $t = x - \sigma$, it is easy to see that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\sigma)^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = 1; \tag{4.14}$$

cf. (3.33). From (4.13) and (4.14), we conclude that

$$E[Y] = e^{\mu+\frac{\sigma^2}{2}}.$$

To compute the variance of Y, recall from Lemma 3.5 that

$$\mathrm{var}(Y) = E[Y^2] - (E[Y])^2.$$

The subsequent computations are similar to those for computing $E[Y]$, and require once again the completion of the square in the exponent of the exponential function. Recall that $Y = e^{\mu+\sigma Z}$ and $E[Y] = e^{\mu+\frac{\sigma^2}{2}}$. Then,

$$\begin{aligned}
\mathrm{var}(Y) &= E[Y^2] - (E[Y])^2 = E[e^{2\mu+2\sigma Z}] - \left(e^{\mu+\frac{\sigma^2}{2}}\right)^2 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{2\mu+2\sigma y}\, e^{-\frac{y^2}{2}}\, dy - e^{2\mu+\sigma^2} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{2\mu+2\sigma^2-\frac{(y-2\sigma)^2}{2}}\, dy - e^{2\mu+\sigma^2} \\
&= e^{2\mu+2\sigma^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-2\sigma)^2}{2}}\, dy - e^{2\mu+\sigma^2} \\
&= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right).
\end{aligned}$$

Here, we used the fact that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-2\sigma)^2}{2}}\, dy = 1,$$

which follows from (3.33) by using the substitution $x = y - 2\sigma$. □

4.3 Independent random variables

Finding the joint distribution function of random variables is often challenging. However, the joint distribution function of independent random variables is the product of the density functions of the respective random variables.

Definition 4.2. *Two random variables X_1 and X_2 over the same probability space are independent if and only if*

$$P((a_1 \le X_1 \le b_1) \cap (a_2 \le X_2 \le b_2)) = P(a_1 \le X_1 \le b_1)\; P(a_2 \le X_2 \le b_2),$$

for any real numbers $a_1 \le b_1$ and $a_2 \le b_2$.

The following result can be regarded as an equivalent definition of two independent random variables:

Lemma 4.4. *Two random variables X_1 and X_2 over the same probability space are independent if and only if*

$$P((a_1 \le X_1 \le b_1) \mid (a_2 \le X_2 \le b_2)) = P(a_1 \le X_1 \le b_1), \tag{4.15}$$

for any real numbers $a_1 \le b_1$ and $a_2 \le b_2$ such that $P(a_2 \le X_2 \le b_2) > 0$.

The notation $P(A|B)$ from (4.15) is the conditional probability of the event A happening provided that event B has occured (with probability $P(B) > 0$), and is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Definition 4.3. *Let X_1 and X_2 be two random variables over the same probability space. The joint distribution function $f : \mathbb{R}^2 \to [0, \infty)$ of X_1 and X_2 is defined by*

$$P((a_1 \leq X_1 \leq b_1) \cap (a_2 \leq X_2 \leq b_2)) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2)\, dx_1\, dx_2, \quad (4.16)$$

for any real numbers $a_1 \leq b_1$ and $a_2 \leq b_2$.

Lemma 4.5. *Two random variables X_1 and X_2 over the same probability space are independent if and only if the joint distribution function $f(x_1, x_2)$ of X_1 and X_2 is*

$$f(x_1, x_2) = f_1(x_1) f_2(x_2), \quad (4.17)$$

where $f_1(x)$ and $f_2(x)$ are the probability density functions of X_1 and X_2, respectively.

Lemma 4.6. *Let X_1 and X_2 be two independent random variables over the same probability space, and let $h_1 : \mathbb{R} \to \mathbb{R}$ and $h_2 : \mathbb{R} \to \mathbb{R}$ be continuous functions. Then*

$$E[h_1(X_1)\, h_2(X_2)] = E[h_1(X_1)]\, E[h_2(X_2)].$$

Proof. Let $f_1(x)$ and $f_2(x)$ be the probability density functions of X_1 and X_2, respectively, and let $f(x_1, x_2)$ be the joint distribution function of X_1 and X_2. A result analogous to Lemma 3.4 holds for the random variables X_1 and X_2 and for the function $h_1(X_1)\, h_2(X_2)$, i.e.,

$$E[h_1(X_1) h_2(X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(x_1) h_2(x_2) f(x_1, x_2)\, dx_1\, dx_2. \quad (4.18)$$

Recall from (4.17) that $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, since X_1 and X_2 are independent. Then, from (4.18), and using Lemma 3.4, we conclude that

$$\begin{aligned}
E[h_1(X_1) h_2(X_2)] &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h_1(x_1) h_2(x_2) f_1(x_1) f_2(x_2)\, dx_1 \right) dx_2 \\
&= \int_{-\infty}^{\infty} h_2(x_2) f_2(x_2) \left(\int_{-\infty}^{\infty} h_1(x_1) f_1(x_1)\, dx_1 \right) dx_2 \\
&= \left(\int_{-\infty}^{\infty} h_1(x_1) f_1(x_1)\, dx_1 \right) \left(\int_{-\infty}^{\infty} h_2(x_2) f_2(x_2)\, dx_2 \right) \\
&= E[h_1(X_1)]\, E[h_2(X_2)],
\end{aligned}$$

which is what we wanted to prove. □

Lemma 4.7. *Let X_1 and X_2 be two independent random variables over the same probability space. Then*

$$cov(X_1, X_2) = corr(X_1, X_2) = 0, \quad (4.19)$$

and therefore

$$var(X_1 + X_2) = var(X_1) + var(X_2). \quad (4.20)$$

Proof. From Lemma 4.6, we find that $E[X_1 X_2] = E[X_1]\, E[X_2]$. Then, from Lemma 3.6, it follows that

$$\text{cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]\, E[X_2] = 0,$$

and, from definition (3.18), it follows that $\text{corr}(X_1, X_2) = 0$.

Using Lemma 3.7, we conclude that

$$var(X_1 + X_2) = var(X_1) + 2\text{cov}(X_1, X_2) + var(X_2) = var(X_1) + var(X_2).$$

□

The next two results are given without proofs, which are technical and beyond the scope of this book.

Lemma 4.8. *Let X_1 and X_2 be independent random variables over the same probability space, and let $h_1 : \mathbb{R} \to \mathbb{R}$ and $h_2 : \mathbb{R} \to \mathbb{R}$ be continuous functions. Then the random variables $h_1(X_1)$ and $h_2(X_2)$ are also independent.*

Theorem 4.1. *Let X_1 and X_2 be independent random variables over the same probability space, with probability density functions $f_1(x)$ and $f_2(x)$, respectively. If X_1 and X_2 are independent, then the probability density function of the random variable $X_1 + X_2$ is the convolution function $(f_1 * f_2)(x)$, i.e.,*

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(z) f_2(x - z)\, dz = \int_{-\infty}^{\infty} f_1(x - z) f_2(x)\, dz.$$

While it is not necessarily true that the sum of two normal variables is a normal variable, that is the case if the two normal variables are *independent.*

Theorem 4.2. *Let X_1 and X_2 be independent normal random variables of mean and variance μ_1 and σ_1^2, and μ_2 and σ_2^2, respectively. Then $X_1 + X_2$ is a normal variable of mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.*

Proof. The density functions $f_1(x)$ and $f_2(x)$ of X_1 and X_2, respectively, are

$$f_1(x) = \frac{1}{|\sigma_1|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right); \tag{4.21}$$

$$f_2(x) = \frac{1}{|\sigma_2|\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right); \tag{4.22}$$

cf. Lemma 3.14.

Since X_1 and X_2 are independent, it follows from Theorem 4.1 that the probability density function $f(z)$ of $X_1 + X_2$ is the convolution of the probability densities of X_1 and X_2, i.e.,

$$\begin{aligned} f(z) &= (f_1 * f_2)(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x)\,dx \\ &= \frac{1}{2\pi\,|\sigma_1\sigma_2|} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(z-x-\mu_2)^2}{2\sigma_2^2}\right) dx. \end{aligned}$$

By completing the square, we can write the term

$$-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(z-x-\mu_2)^2}{2\sigma_2^2}$$

as

$$-\frac{\left(x - (\mu_1\sigma_2^2 + (z-\mu_2)\sigma_1^2)\right)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)} - \frac{(z - (\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}.$$

Therefore, the density function $f(z)$ of $X_1 + X_2$ can be written as

$$\begin{aligned} f(z) &= \frac{1}{2\pi\,|\sigma_1\sigma_2|} \exp\left(-\frac{(z-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right) \\ &\quad \int_{-\infty}^{\infty} \exp\left(-\frac{\left(x - (\mu_1\sigma_2^2 + (z-\mu_2)\sigma_1^2)\right)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}\right) dx. \end{aligned}$$

For the last integral, we use the substitution

$$y = \frac{x - (\mu_1\sigma_2^2 + (z-\mu_2)\sigma_1^2)}{|\sigma_1\sigma_2|/\sqrt{\sigma_1^2+\sigma_2^2}}.$$

The function under the integration sign becomes $\exp\left(-\frac{y^2}{2}\right)$. Since

$$dx = \frac{|\sigma_1\sigma_2|}{\sqrt{\sigma_1^2+\sigma_2^2}}\,dy,$$

and using the fact that

$$\int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = \sqrt{2\pi},$$

see (3.33), we conclude that

$$\int_{-\infty}^{\infty} \exp\left(-\frac{\left(x - (\mu_1\sigma_2^2 + (z-\mu_2)\sigma_1^2)\right)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}\right) dx = \sqrt{2\pi}\,\frac{|\sigma_1\sigma_2|}{\sqrt{\sigma_1^2+\sigma_2^2}}.$$

Therefore,

$$\begin{aligned} f(z) &= \frac{1}{2\pi\,|\sigma_1\sigma_2|} \exp\left(-\frac{(z-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right) \frac{\sqrt{2\pi}\,|\sigma_1\sigma_2|}{\sqrt{\sigma_1^2+\sigma_2^2}} \\ &= \frac{1}{\sqrt{2\pi\,(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(z-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right). \end{aligned}$$

We proved that the density function of $X_1 + X_2$ is

$$f(z) = \frac{1}{\sqrt{2\pi\,(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(z-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right).$$

From Lemma 3.14, we conclude that $X_1 + X_2$ is a normal variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. □

If two normal variables are not independent, then their sum need not be a normal variable. Given our definition (3.44), this requires a brief clarification. If X_1 is a normal variable of mean μ_1 and variance σ_1^2, then

$$X_1 = \mu_1 + \sigma_1 Z. \tag{4.23}$$

If X_2 is another normal variable of mean μ_2 and variance σ_2^2, then X_2 can also be written as the sum of μ_2 and σ_2 times the standard normal variable, but this new standard normal variable, although with the same probability distribution as Z from (4.23), is a different standard normal. To avoid confusions, we denote this new standard normal by Z_2 and the standard normal variable Z from (4.23) by Z_1. Then,

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Z_1; \\ X_2 &= \mu_2 + \sigma_2 Z_2. \end{aligned}$$

Therefore, it is not true that $X_1 + X_2 = \mu_1 + \mu_2 + (\sigma_1 + \sigma_2)Z$, which would have meant that $X_1 + X_2$ is normal. As a matter of fact, from Theorem 4.2 we know that, if X_1 and X_2 are independent, then

$$X_1 + X_2 = \mu_1 + \mu_2 + \sqrt{\sigma_1^2 + \sigma_2^2}\,Z,$$

since $X_1 + X_2$ is a normal variable of mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

The sum of two lognormal variables, independent or not, is not lognormal. However, the product of two independent lognormal variables is lognormal.

Theorem 4.3. *Let Y_1 and Y_2 be independent lognormal random variables with parameters μ_1 and σ_1 and μ_2 and σ_2, respectively. Then $\ln(Y_1\ Y_2)$ is a lognormal variable with parameters $\mu_1 + \mu_2$ and $\sqrt{\sigma_1^2 + \sigma_2^2}$.*

Proof. Note that $\ln(Y_1) = X_1 = \mu_1 + \sigma_1 Z_1$ and $\ln(Y_2) = X_2 = \mu_2 + \sigma_2 Z_2$, where Z_1 and Z_2 are two independent standard normal variables. Since Y_1 and Y_2 are independent, we find, from Lemma 4.8, that the corresponding normal variables X_1 and X_2 are also independent. From Theorem 4.2, it follows that $X_1 + X_2$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$, i.e.,

$$X_1 + X_2 = \mu_1 + \mu_2 + \sqrt{\sigma_1^2 + \sigma_2^2}\ Z.$$

Therefore,

$$\ln(Y_1\ Y_2) = \ln(Y_1) + \ln(Y_2) = X_1 + X_2 = (\mu_1 + \mu_2) + \sqrt{\sigma_1^2 + \sigma_2^2}\ Z,$$

which shows that $\ln(Y_1\ Y_2)$ is a lognormal random variable with parameters mean $\mu_1 + \mu_2$ and $\sqrt{\sigma_1^2 + \sigma_2^2}$. □

4.4 Approximating sums of lognormal variables

The product of two independent lognormal variables is lognormal variable, while their sum is not. However, approximating the sum of two lognormal variables by a lognormal variable is useful in practice, e.g., for modeling derivative securities involving two or more assets.

For example, the payoff of a basket option depends on the sum of the prices of several assets[1]. One way to price a basket option is to approximate the value of the underlying asset of the option, i.e., of the sum of the prices of the assets, with a lognormal variable, and use the Black–Scholes framework. The parameters of the lognormal variable are chosen such that the first two moments of the sum of the asset prices will match the expected value and the variance of the lognormal variable.

[1]The underlying asset of a basket option is, in general, a weighted average of the prices of several assets, and not necessarily their sum.

Assume that the prices Y_1 and Y_2 of the two assets are lognormal random variables with parameters μ_1 and σ_1^2 and μ_2 and σ_2^2, respectively. We are looking for a lognormal variable Y such that

$$Y_1 + Y_2 \approx Y.$$

To make this precise, let $\ln(Y) = \mu + \sigma Z$. We look for μ and σ such that the expected value and the variance of Y and of $Y_1 + Y_2$ are equal, i.e.,

$$E[Y] = E[Y_1 + Y_2]; \tag{4.24}$$

$$\mathrm{var}(Y) = \mathrm{var}(Y_1 + Y_2). \tag{4.25}$$

From Lemma 4.3, it is easy to see that

$$E[Y] = e^{\mu + \frac{\sigma^2}{2}}; \quad \mathrm{var}(Y) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right);$$

$$E[Y_i] = e^{\mu_i + \frac{\sigma_i^2}{2}}; \quad \mathrm{var}(Y_i) = e^{2\mu_i+\sigma_i^2}\left(e^{\sigma_i^2} - 1\right), \quad i = 1:2.$$

From Lemma 3.3 and Lemma 3.7, we find that

$$E[Y_1 + Y_2] = E[Y_1] + E[Y_2] = e^{\mu_1 + \frac{\sigma_1^2}{2}} + e^{\mu_2 + \frac{\sigma_2^2}{2}};$$

$$\mathrm{var}(Y_1 + Y_2) = \mathrm{var}(Y_1) + 2\rho_{1,2}\sqrt{\mathrm{var}(Y_1)\ \mathrm{var}(Y_2)} + \mathrm{var}(Y_2),$$

where $\rho_{1,2} = \mathrm{corr}(X, Y)$ is the correlation between Y_1 and Y_2.

We can now write (4.24) and (4.25) as

$$e^{\mu + \frac{\sigma^2}{2}} = e^{\mu_1 + \frac{\sigma_1^2}{2}} + e^{\mu_2 + \frac{\sigma_2^2}{2}}; \tag{4.26}$$

$$e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{2\mu_1+\sigma_1^2}\left(e^{\sigma_1^2} - 1\right) + e^{2\mu_2+\sigma_2^2}\left(e^{\sigma_2^2} - 1\right) \tag{4.27}$$

$$+2\rho_{1,2}\sqrt{e^{2\mu_1+2\mu_2+\sigma_1^2+\sigma_2^2}\left(e^{\sigma_1^2} - 1\right)\left(e^{\sigma_2^2} - 1\right)}.$$

From (4.26), we find that

$$e^{2\mu+\sigma^2} = \left(e^{\mu_1 + \frac{\sigma_1^2}{2}} + e^{\mu_2 + \frac{\sigma_2^2}{2}}\right)^2,$$

and, by substituting in (4.27), we obtain an equation depending only on σ:

$$\left(e^{\mu_1 + \frac{\sigma_1^2}{2}} + e^{\mu_2 + \frac{\sigma_2^2}{2}}\right)^2\left(e^{\sigma^2} - 1\right) = e^{2\mu_1+\sigma_1^2}\left(e^{\sigma_1^2} - 1\right) + e^{2\mu_2+\sigma_2^2}\left(e^{\sigma_2^2} - 1\right)$$

$$+2\rho_{1,2}\sqrt{e^{2\mu_1+2\mu_2+\sigma_1^2+\sigma_2^2}\left(e^{\sigma_1^2} - 1\right)\left(e^{\sigma_2^2} - 1\right)}.$$

This equation can be written in a simpler form as

$$\left(e^{\mu_1+\frac{\sigma_1^2}{2}}+e^{\mu_2+\frac{\sigma_2^2}{2}}\right)^2 e^{\sigma^2} = e^{2\mu_1+2\sigma_1^2} + e^{2\mu_2+2\sigma_2^2}$$
$$+2e^{\mu_1+\mu_2+\frac{\sigma_1^2+\sigma_2^2}{2}}\left(1+\rho_{1,2}\sqrt{\left(e^{\sigma_1^2}-1\right)\left(e^{\sigma_2^2}-1\right)}\right)$$

and solving for σ we obtain

$$\sigma = \left(\ln\left(\frac{h(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho_{1,2})}{\left(e^{\mu_1+\sigma_1^2/2} + e^{\mu_2+\sigma_2^2/2}\right)^2}\right)\right)^{\frac{1}{2}}, \tag{4.28}$$

where

$$h(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho_{1,2}) = e^{2\mu_1+2\sigma_1^2} + e^{2\mu_2+2\sigma_2^2}$$
$$+2e^{\mu_1+\mu_2+\frac{\sigma_1^2+\sigma_2^2}{2}}\left(1+\rho_{1,2}\sqrt{\left(e^{\sigma_1^2}-1\right)\left(e^{\sigma_2^2}-1\right)}\right).$$

The value of μ can be obtained from (4.26) and (4.28) as follows:

$$\mu = \ln\left(e^{\mu_1+\frac{\sigma_1^2}{2}} + e^{\mu_2+\frac{\sigma_2^2}{2}}\right) - \frac{\sigma^2}{2}.$$

4.5 Power series

In this section, we discuss the convergence and smoothness properties of power series. We note that Taylor series expansions are special cases of power series; see section 5.3.

Definition 4.4. *The power series*

$$S(x) = \sum_{k=0}^{\infty} a_k(x-a)^k \tag{4.29}$$

is defined as

$$S(x) = \lim_{n\to\infty} \sum_{k=0}^{n} a_k(x-a)^k \tag{4.30}$$

for all the values $x \in \mathbb{R}$ such that the limit $\lim_{n\to\infty} \sum_{k=0}^n a_k(x-a)^k$ exists and is finite.

Example: A simple example of power series is the geometric series

$$\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty} \sum_{k=0}^{n} x^k, \tag{4.31}$$

which is convergent if and only if $|x| < 1$.

Answer: Note that the geometric series can be written in the general form (4.30) with $a = 0$ and $a_k = 1$, $\forall\ k \geq 0$.

Recall from (21) that, for any $x \neq 1$,

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}. \tag{4.32}$$

Therefore, if $|x| < 1$, the geometric series (4.31) is convergent and

$$\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty}\sum_{k=0}^{n} x^k = \lim_{n\to\infty}\frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}.$$

The geometric series is not convergent if $|x| \geq 1$.

If $|x| > 1$, then $\left|x^{n+1}\right| \to \infty$ when $n \to \infty$, and the partial sums $\sum_{k=0}^n x^k$ do not converge; cf. (4.32).

If $x = 1$, the partial sum $\sum_{k=0}^n x^k$ is equal to n, and therefore does not converge as $n \to \infty$ as required in (4.31).

If $x = -1$, the partial sum $\sum_{k=0}^n (-1)^k$ is equal to 1, if n is even, and to 0, if n is odd. Therefore, the limit $\lim_{n\to\infty}\sum_{k=0}^n x^k$ does not exist. □

To present convergence results for arbitrary power series, we first define the radius of convergence of a series. Recall that, if $(x_k)_{k=1:\infty}$ is a sequence of real numbers, then $\limsup_{k\to\infty} x_k$ is the highest value l for which there exists a subsequence of $(x_k)_{k=1:\infty}$ that converges to l.

Definition 4.5. *The radius of convergence R of the power series $S(x) = \sum_{k=0}^\infty a_k(x-a)^k$ is defined as*

$$R = \frac{1}{\limsup_{k\to\infty}|a_k|^{1/k}}. \tag{4.33}$$

If $\lim_{k\to\infty}|a_k|^{1/k}$ exists, then $\limsup_{k\to\infty}|a_k|^{1/k} = \lim_{k\to\infty}|a_k|^{1/k}$. The following result follows directly from definition (4.33):

Lemma 4.9. *Let $S(x) = \sum_{k=0}^\infty a_k(x-a)^k$ be a power series and let R be the radius of convergence of $S(x)$. If $\lim_{k\to\infty}|a_k|^{1/k}$ exists, then*

$$R = \frac{1}{\lim_{k\to\infty}|a_k|^{1/k}}. \tag{4.34}$$

Theorem 4.4. *Let $S(x) = \sum_{k=0}^{\infty} a_k(x-a)^k$ be a power series and let R be the radius of convergence of $S(x)$. Then,*

- *if $R = 0$, the series $S(x)$ is only convergent at the point $x = a$;*
- *if $R = \infty$, the series $S(x)$ is (absolutely) convergent for all $x \in \mathbb{R}$;*
- *if $0 < R < \infty$, then the series $S(x)$ is (absolutely) convergent if $|x-a| < R$ and is not convergent if $|x-a| > R$; if $|x-a| = R$, i.e., if either $x-a = -R$, or $x - a = R$, the series $S(x)$ could be convergent or divergent at the point x.*

Example: Find the radius of convergence of

$$\sum_{k=1}^{\infty} \frac{x^k}{k^2}.$$

Answer: It is easy to see that $\sum_{k=1}^{\infty} \frac{x^k}{k^2} = \sum_{k=1}^{\infty} a_k x^k$, with $a_k = \frac{1}{k^2}$, $k \geq 1$. Note that

$$\lim_{k\to\infty} |a_k|^{1/k} = \lim_{k\to\infty} \left(\frac{1}{k^2}\right)^{1/k} = \lim_{k\to\infty} \frac{1}{(k^{1/k})^2} = 1,$$

since $\lim_{k\to\infty} k^{1/k} = 1$; cf. (1.29). From (4.34), it follows that the radius of convergence of the series is $R = 1$. We conclude from Theorem 4.4 that the series is convergent if $|x| < 1$, and not convergent if $|x| > 1$.

If $x = -1$, the series becomes $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}$. Since the terms $\frac{(-1)^k}{k^2}$ have alternating signs and decrease in absolute value to 0, the series is convergent.

If $x = 1$, the series becomes $\sum_{k=1}^{\infty} \frac{1}{k^2}$, which is convergent[2]. □

The smoothness properties of power series are presented in the following theorem:

Theorem 4.5. *Let $S(x) = \sum_{k=0}^{\infty} a_k(x-a)^k$ be a power series and let $R > 0$ be the radius of convergence of $S(x)$.*

(i) The function $S(x)$ is differentiable on the interval $(a - R,\ a + R)$, i.e, for $|x - a| < R$, and $S'(x)$ is given by

$$S'(x) = \sum_{k=1}^{\infty} k a_k (x-a)^{k-1}, \quad \forall\ x \in (a - R,\ a + R).$$

(ii) The function $S(x) = \sum_{k=0}^{\infty} a_k(x-a)^k$ is infinitely many times differentiable on the interval $(a - R, a + R)$, and the coefficients a_k are given by

$$a_k = \frac{S^{(k)}(a)}{k!}, \quad \forall\ k \geq 0.$$

[2]It can be shown that $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$.

4.5.1 Stirling's formula

Let $k! = 1 \cdot 2 \cdot \ldots \cdot (k-1) \cdot k$ be the factorial of the positive integer k. Note that $k!$ is a large number even for k small; for example, $10! = 3,628,800$.

Stirling's formula gives the order of magnitude for $k!$ when k is large, i.e.,

$$k! \simeq \left(\frac{k}{e}\right)^k \sqrt{2\pi k}. \tag{4.35}$$

More rigorously, Stirling's formula says that

$$\lim_{k\to\infty} \frac{k!}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k}} = 1. \tag{4.36}$$

In section 5.3, we will use the following weaker version of (4.36) to study the convergence of Taylor series expansions:

$$\lim_{k\to\infty} \frac{(k!)^{1/k}}{k} = \frac{1}{e}. \tag{4.37}$$

To prove (4.37), note that, if (4.36) holds true, then

$$\lim_{k\to\infty} \left(\frac{k!}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k}} \right)^{1/k} = 1$$

and therefore

$$1 = \lim_{k\to\infty} \frac{(k!)^{1/k}}{\frac{k}{e}\, (\sqrt{2\pi k})^{1/k}} = e \lim_{k\to\infty} \frac{(k!)^{1/k}}{k}, \tag{4.38}$$

since we can use (1.29) and (1.30) to show that

$$\lim_{k\to\infty} (\sqrt{2\pi k})^{1/k} = \left(\lim_{k\to\infty} (\sqrt{2\pi})^{1/k}\right) \cdot \left(\lim_{k\to\infty} \sqrt{k^{1/k}}\right) = 1.$$

The limit (4.38) is equivalent to (4.37), which is what we wanted to show.

FINANCIAL APPLICATIONS

A lognormal model for the evolution of asset prices.

The risk–neutral derivation of the Black–Scholes formula.

Financial interpretation of $N(d_1)$ and $N(d_2)$ from the Black–Scholes formula.

The probability that a plain vanilla European option expires in the money.

4.6 A lognormal model for the evolution of asset prices

To price derivative securities, we need to make assumptions on the evolution of the price of the underlying assets. The most elementary assumption that does not contradict market data is that the price of the asset has lognormal distribution; see (3.52). The fact that, under the lognormal assumption, closed formulas for pricing certain options can be derived, e.g., the Black–Scholes formulas from section 3.5, is most welcome and explains the wide use of this model.

To justify the lognormal assumption, note that the rate of return of an asset between times t and $t + \delta t$, i.e., over an infinitesimal time period δt, is

$$\frac{S(t+\delta t) - S(t)}{S(t)\ \delta t}. \tag{4.39}$$

If the rate of return (4.39) were constant, i.e., if there existed a parameter μ such that

$$\frac{S(t+\delta t) - S(t)}{\delta t\ S(t)} \ = \ \mu, \tag{4.40}$$

then the price $S(t)$ of the underlying asset would be an exponential function: When $\delta t \to 0$, formula (4.40) becomes the ODE

$$\frac{dS}{dt} \ = \ \mu S, \ \ \forall\, 0 < t < T,$$

with initial condition at $t = 0$ given by $S(0)$, the price of the asset at time 0. The solution of the ODE is

$$S(t) \ = \ S(0)\ e^{\mu t}.$$

Since market data shows that asset prices do not grow exponentially, the constant rate of return assumption cannot be correct.

A more realistic model is to assume that the rate of return (4.39) has a random oscillation around its mean. This oscillation is assumed to be normally distributed, with mean 0 and standard deviation on the order of $\frac{1}{\sqrt{\delta t}}$. (Any order of magnitude other than $\frac{1}{\sqrt{\delta t}}$ would render the asset price either infinitely large or infinitely small.) In other words, we consider the following model for $S(t)$:

$$\frac{S(t+\delta t) - S(t)}{S(t)\ \delta t} \ = \ \mu \ + \ \sigma \frac{1}{\sqrt{\delta t}} Z,$$

which is equivalent to

$$S(t+\delta t) - S(t) \ = \ \mu S(t)\delta t \ + \ \sigma S(t)\sqrt{\delta t} Z. \tag{4.41}$$

The parameters μ and σ are called the drift and the volatility of the underlying asset, respectively, and represent the expected value and the standard deviation of the returns of the asset. Based on (4.41), we require $S(t)$ to satisfy the following stochastic differential equation (SDE):

$$dS \ = \ \mu S dt \ + \ \sigma S dX, \tag{4.42}$$

with initial condition at $t = 0$ given by $S(0)$; here, $X(t)$, $t \geq 0$, is a Weiner process $X(t)$. If the asset pays dividends continuously at rate q, the SDE (4.42) becomes

$$dS \ = \ (\mu - q) S dt \ + \ \sigma S dX. \tag{4.43}$$

We will not go into any further details on SDEs. For our purposes we only need the fact that $S(t)$ is the solution of the SDE (4.43) for $t \geq 0$ if and only if $S(t)$ satisfies the following lognormal model:

$$\ln\left(\frac{S(t_2)}{S(t_1)}\right) \ = \ \left(\mu - q - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma\sqrt{t_2 - t_1} Z, \ \forall\, 0 \leq t_1 < t_2, \tag{4.44}$$

where Z is the standard normal variable.

4.7 Risk–neutral derivation of the Black–Scholes formula

Risk–neutral pricing refers to pricing derivative securities as discounted expected values of their payoffs at maturity, under the assumption that the underlying asset has lognormal distribution with drift μ equal to the constant risk-free rate r. In other words,

$$V(0) \ = \ e^{-rT}\ E_{RN}[V(T)], \tag{4.45}$$

where the expected value in (4.45) is computed with respect to $S(T)$ given by (4.44) for $\mu = r$, $t_1 = 0$ and $t_2 = T$, i.e.,

$$\ln\left(\frac{S(T)}{S(0)}\right) \ = \ \left(r - q - \frac{\sigma^2}{2}\right) T \ + \ \sigma\sqrt{T} Z, \tag{4.46}$$

which is equivalent to

$$S(T) \ = \ S(0)\ e^{\left(r-q-\frac{\sigma^2}{2}\right)T \ + \ \sigma\sqrt{T}Z}. \tag{4.47}$$

While risk–neutral pricing does not hold for path–dependent options or American options, it can be used to price plain vanilla European Call and Put options, and is one way to derive the Black–Scholes formulas.

Applying (4.45) to plain vanilla European options, we find that

$$C(0) = e^{-rT} E_{RN}[\max(S(T)-K,0)]; \tag{4.48}$$
$$P(0) = e^{-rT} E_{RN}[\max(K-S(T),0)], \tag{4.49}$$

where the expected value is computed with respect to $S(T)$ given by (4.47).

We derive the Black–Scholes formula for call options using (4.48). The Black–Scholes formula for put options can be obtained similarly from (4.49).

By definition,

$$\max(S(T)-K,0) = \begin{cases} S(T)-K, & \text{if } S(T) \geq K; \\ 0, & \text{otherwise.} \end{cases}$$

From (4.47), it follows that

$$S(T) \geq K \iff Z \geq \frac{\ln\left(\frac{K}{S(0)}\right) - \left(r-q-\frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}} = -d_2,$$

where we used the notation

$$d_2 = \frac{\ln\left(\frac{S(0)}{K}\right) + \left(r-q-\frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}. \tag{4.50}$$

Note that the term d_2 from (4.50) is the same as the term d_2 from the Black–Scholes formula given by (3.56), if $t=0$. Thus,

$$\max(S(T)-K,0) = \begin{cases} S(0)e^{\left(r-q-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z} - K & \text{if } Z \geq -d_2; \\ 0 & \text{if } Z < -d_2. \end{cases} \tag{4.51}$$

Recall that the probability density function of the standard normal variable Z is $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. The pricing formula (4.48) can be regarded as an expectation over Z; cf. (4.51). Then, from Lemma 3.4, it follows that

$$\begin{aligned} C(0) &= e^{-rT}\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty}\left(S(0)\, e^{\left(r-q-\frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}x} - K\right) e^{-\frac{x^2}{2}}\,dx \\ &= \frac{S(0)e^{-rT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{\left(r-q-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}x-\frac{x^2}{2}}dx - \frac{Ke^{-rT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{x^2}{2}}dx \\ &= \frac{S(0)e^{-qT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2}T+\sigma\sqrt{T}x-\frac{x^2}{2}}dx - \frac{Ke^{-rT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{x^2}{2}}dx \\ &= \frac{S(0)e^{-qT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{1}{2}\left(x-\sigma\sqrt{T}\right)^2}dx - \frac{Ke^{-rT}}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{x^2}{2}}dx. \end{aligned} \tag{4.52}$$

We make the substitution $y = x - \sigma\sqrt{T}$ in the first integral from (4.52). The integration limits change from $x=\infty$ to $y=\infty$ and from $x=-d_2$ to $y=-d_2-\sigma\sqrt{T}=-d_1$, where we used the notation

$$d_1 = d_2 + \sigma\sqrt{T} = \frac{\ln\left(\frac{S(0)}{K}\right) + \left(r-q+\frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}. \tag{4.53}$$

Note that the term d_1 from (4.53) is the same as the term d_1 from the Black–Scholes formula given by (3.55) if $t=0$.

Formula (4.52) becomes

$$C(0) = S(0)e^{-qT}\frac{1}{\sqrt{2\pi}}\int_{-d_1}^{\infty} e^{-\frac{y^2}{2}}\,dy - Ke^{-rT}\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-\frac{x^2}{2}}\,dx. \tag{4.54}$$

Let $N(t) = P(Z \leq t)$ be the cumulative distribution of Z. Recall from Lemma 3.12 that $1-N(-a) = N(a)$, for any a. Then,

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}\int_{-a}^{\infty} e^{-\frac{x^2}{2}}\,dx &= P(Z \geq -a) = 1 - P(Z \leq -a) \\ &= 1 - N(-a) = N(a), \quad \forall\, a \in \mathbb{R}. \end{aligned} \tag{4.55}$$

From (4.54) and using (4.55) for both $a=d_1$ and $a=d_2$, we conclude that

$$C(0) = S(0)e^{-qT}N(d_1) - Ke^{-rT}N(d_2).$$

This is exactly the Black–Scholes formula (3.53) obtained by setting $t=0$.

4.8 Computing the probability that a European option expires in the money

Consider a plain vanilla European call option at time 0 on an underlying asset with spot price having a lognormal distribution with drift μ and volatility σ, and paying dividends continuously at rate q, i.e.,

$$\ln\left(\frac{S(t_2)}{S(t_1)}\right) = \left(\mu - q - \frac{\sigma^2}{2}\right)(t_2-t_1) + \sigma\sqrt{t_2-t_1}Z, \ \forall\, 0 \leq t_1 < t_2, \tag{4.56}$$

where Z is the standard normal variable. Let K and T be the strike and maturity of the option, respectively, and assume that the risk-free interest rates are constant and equal to r until the maturity of the option.

The probability that the call option will expire in the money and will be exercised is equal to the probability that the spot price at maturity is higher than the strike price, i.e., to $P(S(T) > K)$.

From (4.56) for $t_1 = 0$ and $t_2 = T$, we find that

$$\ln\left(\frac{S(T)}{S(0)}\right) = \left(\mu - q - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z. \quad (4.57)$$

Using the lognormal assumption (4.57), we obtain that

$$\begin{aligned} P(S(T) > K) &= P\left(\frac{S(T)}{S(0)} > \frac{K}{S(0)}\right) = P\left(\ln\left(\frac{S(T)}{S(0)}\right) > \ln\left(\frac{K}{S(0)}\right)\right) \\ &= P\left(\left(\mu - q - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z > \ln\left(\frac{K}{S(0)}\right)\right) \\ &= P\left(Z > \frac{\ln\left(\frac{K}{S(0)}\right) - \left(\mu - q - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right) \\ &= P(Z > a), \end{aligned}$$

where

$$a = \frac{\ln\left(\frac{K}{S(0)}\right) - \left(\mu - q - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

Let $N(t) = P(Z \le t)$ be the cumulative distribution (3.41) of Z. Note that $P(Z = a) = 0$ for any $a \in \mathbb{R}$. Then, from Lemma 3.12, it follows that $P(Z > a) = P(Z < -a)$ and therefore

$$P(S(T) > K) = P(Z > a) = P(Z < -a) = N(-a). \quad (4.58)$$

Let $d_{2,\mu} = -a$, i.e.,

$$d_{2,\mu} = \frac{\ln\left(\frac{S(0)}{K}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}. \quad (4.59)$$

From (4.58) and (4.59), we conclude that the probability that a call option expires in the money is

$$P(S(T) > K) = N(d_{2,\mu}) = N\left(\frac{\ln\left(\frac{S(0)}{K}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right). \quad (4.60)$$

From a practical standpoint, in order to compute the probability that the call option will expire in the money, the cumulative distribution $N(t)$ of the standard normal variable must be estimated numerically. This can be done by using, e.g., the pseudocode from Table 3.1.

The term $d_{2,\mu}$ is very similar to the term d_2 from the Black–Scholes formula. If we let $t = 0$ in (3.56), we obtain that

$$d_2 = \frac{\ln\left(\frac{S(0)}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

If the drift of the underlying asset is equal to the constant risk-free rate, i.e., if $\mu = r$, then $d_{2,\mu} = d_2$ and the probability that the call option expires in the money is $N(d_2)$. A further interpretation of this fact will be given in section 4.9.

The probability that a plain vanilla European put option expires in the money can be obtained numerically from the formula (4.60) corresponding to call options. A put option with strike K and maturity T expires in the money if $S(T) < K$. From (4.60) and Lemma 3.12, we find that

$$P(S(T) < K) = 1 - P(S(T) > K) = 1 - N(d_{2,\mu}) = N(-d_{2,\mu}), \quad (4.61)$$

where $d_{2,\mu}$ is given by (4.59).

4.9 Financial Interpretation of $N(d_1)$ and $N(d_2)$ from the Black–Scholes formula

The Black–Scholes formula for a call option is

$$C = S(0)e^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2),$$

and the Delta of the call is equal to $\Delta(C) = e^{-q(T-t)}N(d_1)$; cf. (3.53) and (3.66). For $q = 0$, we obtain that

$$N(d_1) = \Delta(C).$$

In other words, the term $N(d_1)$ from the Black–Scholes formula for a call option is the Delta of the call if the underlying asset does not pay dividends. Note that this interpretation for $N(d_1)$ is incorrect if the underlying asset pays dividends continuously (since in that case $\Delta(C) = e^{-qT}N(d_1)$).

The term $N(d_2)$ from the Black–Scholes formula represents the risk-neutral probability that the call option expires in the money. Recall from section 4.8 that the probability that a call option on an asset following a log-normal process expires in the money is $N(d_{2,\mu})$, where $d_{2,\mu}$ is given by (4.59). As noted before, if the underlying asset has the same drift as the risk-free interest rate r, i.e., if $\mu = r$, then $d_{2,\mu} = d_2$. Therefore, the risk-neutral

probability that the call option expires in the money is $N(d_2)$. We note that this financial interpretation for $N(d_2)$ also holds if the underlying asset pays dividends continuously.

The Black–Scholes formula (3.54) for a put option states that

$$P \;=\; Ke^{-r(T-t)}N(-d_2) - Se^{-q(T-t)}N(-d_1). \tag{4.62}$$

Recall from (3.67) that $\Delta(P) = -e^{-q(T-t)}N(-d_1)$. If the underlying asset pays no dividends, i.e., if $q = 0$, the term $-N(-d_1)$ is the Delta of the put option. The negative sign represents the fact that the Delta of a long put position is negative; see section 3.7 for more details.

The term $N(-d_2)$ from (4.62) is the risk–neutral probability that the put option expires in the money. A simple way of seeing this is to realize that a put expires in the money if and only if the corresponding call with the same strike and maturity expire out of the money. Since the risk–neutral probability that the call option expires in the money is $N(d_2)$, we find, using (3.43), that the risk–neutral probability that the put option expires in the money is $1 - N(d_2) = N(-d_2)$.

4.10 References

A brief treatment of the lognormal model complete with financial insights can be found in Wilmott et al. [35]. The same book contains an intuitive explanation of risk–neutral pricing, as well as the financial interpretation of the terms from the Black–Scholes formula.

4.11 Exercises

1. Let $X_1 = Z$ and $X_2 = -Z$ be two independent random variables, where Z is the standard normal variable. Show that $X_1 + X_2$ is a normal variable of mean 0 and variance 2, i.e., $X_1 + X_2 = \sqrt{2}Z$.

2. Assume that the normal random variables $X_1, X_2, \ldots, X_n$ of mean μ and variance σ^2 are uncorrelated, i.e, $\mathrm{cov}(X_i, X_j) = 0$, for all $1 \le i \ne j \le n$. (This happens, e.g., if $X_1, X_2, \ldots, X_n$ are independent.) If $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the average of the variables X_i, $i = 1 : n$, show that

$$E[S_n] \;=\; \mu \quad \text{and} \quad \mathrm{var}(S_n) \;=\; \frac{\sigma^2}{n}.$$

3. Assume we have a one period binomial model for the evolution of the price of an underlying asset between time t and time $t + \delta t$:

   If $S(t)$ is the price of the asset at time t, then the price $S(t+\delta t)$ of the asset at time $t+\delta t$ will be either $S(t)u$, with (risk–neutral) probability p, or $S(t)d$, with probability $1-p$. Assume that $u > 1$ and $d < 1$.

   Show that

$$E_{RN}[S(t+\delta t)] \;=\; (pu \;+\; (1-p)d)\; S(t); \tag{4.63}$$
$$E_{RN}[S^2(t+\delta t)] \;=\; (pu^2 \;+\; (1-p)d^2)\; S^2(t). \tag{4.64}$$

4. If the price $S(t)$ of a non–dividend paying asset has lognormal distribution with drift $\mu = r$ and volatility σ, then,

$$\ln\left(\frac{S(t+\delta t)}{S(t)}\right) \;=\; \left(r - \frac{\sigma^2}{2}\right)\delta t \;+\; \sigma\sqrt{\delta t}Z.$$

   Show that

$$E_{RN}[S(t+\delta t)] \;=\; e^{r\delta t}S(t); \tag{4.65}$$
$$E_{RN}[S^2(t+\delta t)] \;=\; e^{(2r+\sigma^2)\delta t}S^2(t). \tag{4.66}$$

5. The results of the previous two exercises can be used to calibrate a binomial tree model to a lognormally distributed process. This means finding the up and down factors u and d, and the risk–neutral probability p (of going up) such that the values of $E_{RN}[S(t+\delta t)]$ and $E_{RN}[S^2(t+\delta t)]$

given by (4.63) and (4.64) coincide with the values (4.65) and (4.66) for the lognormal model.

In other words, we are looking for u, d, and p such that

$$pu + (1-p)d = e^{r\delta t} \tag{4.67}$$
$$pu^2 + (1-p)d^2 = e^{(2r+\sigma^2)\delta t} \tag{4.68}$$

Since there are two constraints and three unknowns, the solution will not be unique.

(i) Show that (4.67–4.68) are equivalent to

$$p = \frac{e^{r\delta t} - d}{u-d} \tag{4.69}$$
$$\left(e^{r\delta t} - d\right)\left(u - e^{r\delta t}\right) = e^{2r\delta t}\left(e^{\sigma^2\delta t} - 1\right) \tag{4.70}$$

(ii) Derive the Cox–Ross–Rubinstein parametrization for a binomial tree, by solving (4.69–4.70) with the additional condition that

$$ud = 1.$$

Show that the solution can be written as

$$p = \frac{e^{r\delta t} - d}{u-d}; \quad u = A + \sqrt{A^2-1}; \quad d = A - \sqrt{A^2-1},$$

where

$$A = \frac{1}{2}\left(e^{-r\delta t} + e^{(r+\sigma^2)\delta t}\right).$$

6. Show that the series $\sum_{k=1}^{\infty}\frac{1}{k^2}$ is convergent, while the series $\sum_{k=1}^{\infty}\frac{1}{k}$ and $\sum_{k=2}^{\infty}\frac{1}{k\ln(k)}$ are divergent, i.e., equal to ∞.

   Note: It is known that

   $$\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}$$

   and

   $$\lim_{n\to\infty}\left(\sum_{k=1}^{\infty}\frac{1}{k} - \ln(n)\right) = \gamma,$$

   where $\gamma \approx 0.57721$ is called Euler's constant.

   Hint: Note that

   $$\sum_{k=1}^{n-1}\frac{1}{k+1} < \ln(n) = \int_1^n \frac{1}{x}\,dx = \sum_{k=1}^{n-1}\int_k^{k+1}\frac{1}{x}\,dx < \sum_{k=1}^{n-1}\frac{1}{k},$$

   since $\frac{1}{k+1} < \frac{1}{x} < \frac{1}{k}$, for any $k < x < k+1$. Conclude that

   $$\ln(n) + \frac{1}{n} < \sum_{k=1}^{n}\frac{1}{k} < \ln(n) + 1, \ \forall\, n \geq 1;$$

   For the series $\sum_{k=2}^{\infty}\frac{1}{k\ln(k)}$, use a similar method to find upper and lower bounds for the integral of $\frac{1}{x\ln(x)}$ over the interval $[2, n]$.

7. Find the radius of convergence R of the power series

   $$\sum_{k=2}^{\infty}\frac{x^k}{k\ \ln(k)},$$

   and investigate what happens at the points x where $|x| = R$.

8. Consider a put option with strike 55 and maturity 4 months on a non–dividend paying asset with spot price 60 which follows a lognormal model with drift $\mu = 0.1$ and volatility $\sigma = 0.3$. Assume that the risk-free rate is constant equal to 0.05.

   (i) Find the probability that the put will expire in the money.

   (ii) Find the risk–neutral probability that the put will expire in the money.

   (iii) Compute $N(-d_2)$.

9. (i) Consider an at-the-money call on a non–dividend paying asset; assume the Black-Scholes framework. Show that the Delta of the option is always greater than 0.5.

   (ii) If the underlying pays dividends at the continuous rate q, when is the Delta of an at-the-money call less than 0.5?

   Note: For most cases, the Delta of an at-the-money call option is close to 0.5.

10. Use risk–neutral pricing to price a supershare, i.e., and option that pays $(\max(S(T) - K, 0))^2$ at the maturity of the option. In other words, compute

$$V(0) = e^{-rT} E_{RN}[(\max(S(T) - K, 0))^2],$$

where the expected value is computed with respect to the risk–neutral distribution of the price $S(T)$ of the underlying asset at maturity T, which is assumed to follow a lognormal process with drift r and volatility σ. Assume that the underlying asset pays no dividends, i.e., $q = 0$.

11. If the price of an asset follows a normal process, i.e., $dS = \mu dt + \sigma dX$, then

$$S(t_2) = S(t_1) + \mu(t_2 - t_1) + \sigma\sqrt{t_2 - t_1}\, Z, \quad \forall\, 0 < t_1 < t_2.$$

Assume that the risk free rate is constant and equal to r.

(i) Use risk neutrality to find the value of a call option with strike K and maturity T, i.e., compute

$$C(0) = e^{-rT} E_{RN}[\max(S(T) - K, 0)],$$

where the expected value is computed with respect to $S(T)$ given by

$$S(T) = S(0) + rT + \sigma\sqrt{T}\, Z.$$

(ii) Use the Put–Call parity to find the value of a put option with strike K and maturity T, if the underlying asset follows a normal process as above.

Chapter 5

Taylor's formula and Taylor series. ATM approximation of Black–Scholes formulas.

Taylor's formula for functions of one variable. Derivative and integral forms of the Taylor approximation errors. Convergence of Taylor's formula.

Taylor's formula for multivariable functions.

Taylor series expansions. Convergence properties.

5.1 Taylor's Formula for functions of one variable

Our goal is to approximate a function $f(x)$ around a point a on the real axis, i.e., on an interval $(a-r, a+r)$, by the following polynomial of order n, called the Taylor Polynomial of order n:

$$P_n(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2}f''(a) + \ldots + \frac{(x-a)^n}{n!}f^{(n)}(a), \tag{5.1}$$

which can be written more compactly as

$$P_n(x) = \sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a). \tag{5.2}$$

(Recall that $k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot k$; by convention, $0! = 1$, and $f^{(0)}(a) = f(a)$.)

To decide whether $P_n(x)$ is a good approximation of $f(x)$, the following issues need to be investigated:

1. Convergence: when does $P_n(x)$ converge to $f(x)$ as $n \to \infty$?
2. Order of approximation: how well does the Taylor polynomial $P_n(x)$ approximate $f(x)$?

We present two results that are used to establish the convergence of the Taylor approximations:

Theorem 5.1. ***(Derivative form of the Taylor approximation error.)*** *Let $f(x)$ be a function which is $n+1$ times differentiable, and assume that $f^{(n+1)}(x)$ continuous. Let $f(x) - P_n(x)$ be the approximation error of $f(x)$ by the n-th order Taylor polynomial $P_n(x)$. Then, there exists a point c between a and x such that $f(x) - P_n(x)$ can be expressed as*

$$f(x) - P_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c). \tag{5.3}$$

The proof of this theorem uses Rolle's theorem, and is beyond our scope. Note that, for $n = 0$, Theorem 5.1 becomes the Mean Value Theorem.

Theorem 5.2. ***(Integral form of the Taylor approximation error.)*** *If $f(x)$ is a function which is $n+1$ times differentiable, and if $f^{(n+1)}(x)$ is continuous, then the n-th order Taylor approximation error $f(x) - P_n(x)$ can be expressed as the following integral:*

$$f(x) - P_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\, dt. \tag{5.4}$$

Proof. We give a proof by induction: For $n = 0$, formula (5.4) becomes

$$f(x) - P_0(x) = \int_a^x f'(t)\, dt.$$

Since $P(0) = f(a)$, this is a direct consequence of the Fundamental Theorem of ; cf. Theorem 1.4.

For the induction step, we assume that formula (5.4) holds for n, i.e.,

$$f(x) - P_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\, dt, \tag{5.5}$$

and prove that (5.4) also holds for $n+1$, i.e,

$$f(x) - P_{n+1}(x) = \int_a^x \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t)\, dt. \tag{5.6}$$

Using integration by parts, the term $\int_a^x \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t)\, dt$ can be written as

$$\begin{aligned} & \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+1)}(t)\Big|_{t=a}^{t=x} - \int_a^x \frac{d}{dt}\left(\frac{(x-t)^{n+1}}{(n+1)!}\right) f^{(n+1)}(t)\, dt \\ = & -\frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a) + \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\, dt. \end{aligned}$$

Then, from the induction hypothesis (5.5), it follows that

$$\begin{aligned} \int_a^x \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t)\, dt &= -\frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a) + (f(x) - P_n(x)) \\ &= f(x) - \left(P_n(x) + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a)\right). \end{aligned}$$

From (5.1), it is easy to see that

$$P_{n+1}(x) = P_n(x) + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a).$$

We conclude that

$$\int_a^x \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t)\, dt = f(x) - P_{n+1}(x),$$

which is the same as (5.6). This completes the proof by induction. □

The order of approximation of a function by its Taylor polynomials can be obtained using the derivative form for the Taylor approximation error (5.3).

Theorem 5.3. *Let $f(x)$ be a function which is $n+1$ times differentiable, and assume that $f^{(n+1)}(x)$ is a continuous function. Then, the approximation of $f(x)$ by the n-th order Taylor polynomial is of order $n+1$, i.e.,*

$$\begin{aligned} f(x) = {} & f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2} f''(a) + \ldots + \frac{(x-a)^n}{n!} f^{(n)}(a) \\ & + O\left((x-a)^{n+1}\right), \quad \text{as } x \to a. \end{aligned} \tag{5.7}$$

Proof. Let $P_n(x) = \sum_{k=0}^n \frac{(x-a)^k}{k!} f^{(k)}(a)$ be the n-th order Taylor polynomial corresponding to $f(x)$; cf. (5.1). Then (5.7) can be written as

$$f(x) = P_n(x) + O\left((x-a)^{n+1}\right), \quad \text{as } x \to a. \tag{5.8}$$

Recall from definition (43) that (5.8) is equivalent to proving that

$$\limsup_{x \to a} \frac{|f(x) - P_n(x)|}{|x-a|^{n+1}} < \infty.$$

To show this, consider a fixed interval $[a-r, a+r]$ around a, and let

$$M = \max_{z \in [a-r, a+r]} |f^{(n+1)}(z)|. \tag{5.9}$$

Note that $M < \infty$, since the function $f^{(n+1)}(x)$ is continuous.

Let $x \in [a-r, a+r]$. From the derivative form (5.3) for the Taylor residual, we know that there exists a point c between a and x such that

$$f(x) - P_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c). \tag{5.10}$$

Note that, if c is between a and x and if $x \in [a-r, a+r]$, then $c \in [a-r, a+r]$. From (5.9) and (5.10), we find that

$$\frac{|f(x) - P_n(x)|}{|x-a|^{n+1}} = \frac{|f^{(n+1)}(c)|}{(n+1)!} \le \frac{M}{(n+1)!}.$$

In other words, $\frac{|f(x)-P_n(x)|}{|x-a|^{n+1}}$ is bounded uniformly for all $x \in [a-r, a+r]$, and therefore

$$\limsup_{x \to a} \frac{|f(x) - P_n(x)|}{|x-a|^{n+1}} < \infty,$$

which is what we wanted to prove. □

From Theorem 5.3, we obtain that the linear Taylor approximation

$$f(x) \approx f(a) + (x-a)f'(a) \tag{5.11}$$

is a second order approximation, if $f''(x)$ is a continuous function:

$$f(x) = f(a) + (x-a)f'(a) + O\left((x-a)^2\right), \tag{5.12}$$

as $x \to a$.

Also, the quadratic Taylor approximation

$$f(x) \approx f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2} f''(a) \tag{5.13}$$

is a third order approximation, if $f^{(3)}(x)$ is continuous:

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2} f''(a) + O\left((x-a)^3\right), \tag{5.14}$$

as $x \to a$.

The exponential function, the natural logarithm function, and some simple rational functions appear frequently in financial applications. The following linear and quadratic Taylor approximations are often needed:

Linear Approximations:

$$e^x \approx 1 + x; \tag{5.15}$$
$$\ln(1+x) \approx x; \tag{5.16}$$
$$\ln(1-x) \approx -x; \tag{5.17}$$
$$\frac{1}{1+x} \approx 1 - x; \tag{5.18}$$
$$\frac{1}{1-x} \approx 1 + x. \tag{5.19}$$

Quadratic Approximations:

$$e^x \approx 1 + x + \frac{x^2}{2}; \tag{5.20}$$
$$\ln(1+x) \approx x - \frac{x^2}{2}; \tag{5.21}$$
$$\ln(1-x) \approx -x - \frac{x^2}{2}; \tag{5.22}$$
$$\frac{1}{1+x} \approx 1 - x + x^2; \tag{5.23}$$
$$\frac{1}{1-x} \approx 1 + x + x^2. \tag{5.24}$$

Note that these approximations are accurate only for small values of x.

5.2 Taylor's formula for multivariable functions

Scalar Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables denoted by $x_1, x_2, \ldots, x_n$. Let $x = (x_1, x_2, \ldots, x_n)$. We present first order and second order Taylor expansions of the function $f(x)$, without providing convergence results.

Let $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. The linear Taylor expansion of the function $f(x)$ around the point a,

$$f(x) \approx f(a) + \sum_{i=1}^{n} (x_i - a_i) \frac{\partial f}{\partial x_i}(a), \tag{5.25}$$

is a second order approximation, in the sense that

$$f(x) = f(a) + \sum_{i=1}^{n} (x_i - a_i) \frac{\partial f}{\partial x_i}(a_1, a_2, \ldots, a_n) + O\left(||x-a||^2\right), \tag{5.26}$$

as $x \to a$, if all the partial derivatives of second order of $f(x)$, i.e., $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$, $1 \le i, j \le n$, are continuous. Here, $O\left(||x-a||^2\right) = \sum_{i=1}^{n} O\left(|x_i - a_i|^2\right)$.

The quadratic Taylor expansion of the function $f(x)$ around the point a,

$$f(x) \approx f(a) + \sum_{i=1}^{n} (x_i - a_i) \frac{\partial f}{\partial x_i}(a) + \sum_{1 \le i,j \le n} \frac{(x_i - a_i)(x_j - a_j)}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(a), \tag{5.27}$$

is a third order approximation, in the sense that

$$\begin{aligned} f(x) &= f(a) + \sum_{i=1}^{n}(x_i - a_i)\,\frac{\partial f}{\partial x_i}(a) \\ &+ \sum_{1\le i,j\le n} \frac{(x_i-a_i)(x_j-a_j)}{2}\,\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \\ &+ \sum_{i=1}^{n} O\left(|x_i - a_i|^3\right), \end{aligned} \tag{5.28}$$

as $x \to a$, if all the third order partial derivatives of $f(x)$ are continuous.

The linear and quadratic Taylor expansions (5.25–5.28) can be written using matrix notation as follows: Recall from (1.36) that the gradient $Df(x)$ of the function $f(x)$ is a row vector of size n, i.e.,

$$Df(x) = \left(\frac{\partial f}{\partial x_1}(x) \;\; \frac{\partial f}{\partial x_2}(x) \;\; \ldots \;\; \frac{\partial f}{\partial x_n}(x)\right),$$

and, from (1.37), that the Hessian $D^2 f(x)$ is an $n \times n$ matrix:

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \ldots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

Let

$$x - a = \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{pmatrix}. \tag{5.29}$$

The linear Taylor expansions (5.25) and (5.26) of $f(x)$ around the point a can be written in terms of the gradient $Df(a)$ as

$$f(x) \approx f(a) + Df(a)\,(x-a); \tag{5.30}$$
$$f(x) = f(a) + Df(a)\,(x-a) + O\left(||x-a||^2\right), \tag{5.31}$$

as $x \to a$.

The quadratic Taylor expansions (5.27) and (5.28) of $f(x)$ around the point a can be written in terms of $Df(a)$ and $D^2f(a)$ as

$$f(x) \approx f(a) + Df(a)\,(x-a) + \frac{1}{2}\,(x-a)^t\, D^2 f(a)\,(x-a); \tag{5.32}$$

$$\begin{aligned} f(x) &= f(a) + Df(a)\,(x-a) + \frac{1}{2}\,(x-a)^t\, D^2 f(a)\,(x-a) \\ &+ \sum_{i=1}^{n} O\left(|x_i - a_i|^3\right), \end{aligned} \tag{5.33}$$

as $x \to a$. Note that $(x-a)^t = (x_1 - a_1 \;\; x_2 - a_2 \;\; \ldots \;\; x_n - a_n)$ is a row vector, the transpose of the column vector $x - a$ from (5.29).

Vector Valued Functions

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a vector valued function, given by

$$F(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

where $x = (x_1, x_2, \ldots, x_n)$. Recall from (1.38) that the gradient $DF(x)$ of $F(x)$ is a matrix operator of size $m \times n$ given by

$$DF(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \ldots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}.$$

The linear Taylor expansion of the function $F(x)$ around the point $a \in \mathbb{R}^n$,

$$F(x) \approx F(a) + DF(a)\,(x-a) \tag{5.34}$$

is a second order approximation, i.e.,

$$F(x) = F(a) + DF(a)\,(x-a) + O\left(||x-a||^2\right), \tag{5.35}$$

as $x \to a$, if all the partial order derivatives of order two of the functions $f_k(x)$, $k = 1 : m$, are continuous. As before, $O\left(||x-a||^2\right) = \sum_{i=1}^n O\left(|x_i - a_i|^2\right)$, and $x - a$ is the column vector given by (5.29).

We note that formula (5.34) can be written explicitely as

$$\begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix} \approx \begin{pmatrix} f_1(a) + \sum_{i=1}^n \frac{\partial f_1}{\partial x_i}(a)\,(x_i - a_i) \\ f_2(a) + \sum_{i=1}^n \frac{\partial f_2}{\partial x_i}(a)\,(x_i - a_i) \\ \vdots \\ f_m(a) + \sum_{i=1}^n \frac{\partial f_m}{\partial x_i}(a)\,(x_i - a_i) \end{pmatrix}.$$

Thus, the linear Taylor approximation formula (5.34) for the vector valued function $F(x)$ is obtained by combining the linear Taylor approximation formula (5.25) for each function $f_k(x)$, $k = 1 : m$.

5.2.1 Taylor's formula for functions of two variables

For clarification purposes, we include the formulas for Taylor expansions of both scalar and vector valued functions of two variables in this section.

Scalar Functions

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables. The linear Taylor expansion of $f(x,y)$ around the point $(a,b) \in \mathbb{R}^2$

$$f(x,y) \approx f(a,b) + (x-a)\frac{\partial f}{\partial x}(a,b) + (y-b)\frac{\partial f}{\partial y}(a,b) \tag{5.36}$$

is a second order approximation, i.e.,

$$\begin{aligned} f(x,y) &= f(a,b) + (x-a)\frac{\partial f}{\partial x}(a,b) + (y-b)\frac{\partial f}{\partial y}(a,b) \\ &\quad + O\left(|x-a|^2\right) + O\left(|y-b|^2\right), \end{aligned} \tag{5.37}$$

as $(x,y) \to (a,b)$, if $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, and $\frac{\partial^2 f}{\partial x \partial y}$ are continuous functions.

The quadratic Taylor expansion of $f(x,y)$ around the point $(a,b) \in \mathbb{R}^2$,

$$\begin{aligned} f(x,y) &\approx f(a,b) + (x-a)\frac{\partial f}{\partial x}(a,b) + (y-b)\frac{\partial f}{\partial y}(a,b) \\ &\quad + \frac{(x-a)^2}{2}\frac{\partial^2 f}{\partial x^2}(a,b) + (x-a)(y-b)\frac{\partial^2 f}{\partial x \partial y}(a,b) \\ &\quad + \frac{(y-b)^2}{2}\frac{\partial^2 f}{\partial y^2}(a,b), \end{aligned} \tag{5.38}$$

is a third order approximation, i.e.,

$$\begin{aligned} f(x,y) &= f(a,b) + (x-a)\frac{\partial f}{\partial x}(a,b) + (y-b)\frac{\partial f}{\partial y}(a,b) \\ &\quad + \frac{(x-a)^2}{2}\frac{\partial^2 f}{\partial x^2}(a,b) + (x-a)(y-b)\frac{\partial^2 f}{\partial x \partial y}(a,b) \\ &\quad + \frac{(y-b)^2}{2}\frac{\partial^2 f}{\partial y^2}(a,b) + O\left(|x-a|^3\right) + O\left(|y-b|^3\right), \end{aligned} \tag{5.39}$$

as $(x,y) \to (a,b)$, if all third order derivatives of $f(x,y)$ are continuous.

The matrix forms of the linear Taylor expansions (5.36) and (5.37) are

$$f(x,y) \approx f(a,b) + Df(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix};$$

$$f(x,y) = f(a,b) + Df(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix} + O\left(|x-a|^2\right) + O\left(|y-b|^2\right),$$

as $(x,y) \to (a,b)$, where $Df(a,b) = \left(\frac{\partial f}{\partial x}(a,b) \;\; \frac{\partial f}{\partial y}(a,b)\right)$.

The matrix forms of the quadratic Taylor expansions (5.38) and (5.39) are

$$\begin{aligned} f(x,y) &\approx f(a,b) + Df(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix} \\ &\quad + \frac{1}{2}\begin{pmatrix} x-a \\ y-b \end{pmatrix}^t D^2 f(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix}; \end{aligned} \tag{5.40}$$

$$\begin{aligned} f(x,y) &= f(a,b) + Df(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix} \\ &\quad + \frac{1}{2}\begin{pmatrix} x-a \\ y-b \end{pmatrix}^t D^2 f(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix} \\ &\quad + O\left(|x-a|^3\right) + O\left(|y-b|^3\right), \end{aligned} \tag{5.41}$$

as $(x,y) \to (a,b)$, where

$$D^2 f(a,b) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial x \partial y}(a,b) \\ \frac{\partial^2 f}{\partial y \partial x}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix}.$$

Vector Valued Functions

We conclude by writing linear and quadratic Taylor expansions for a function of two variables taking values in $\mathbb{R}^2$. Let

$$F : \mathbb{R}^2 \to \mathbb{R}^2, \quad \text{given by} \quad F(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix}.$$

Recall from (1.41) that the gradient $DF(x,y)$ of $F(x,y)$ is

$$DF(x,y) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(x,y) & \frac{\partial f_1}{\partial y}(x,y) \\ \frac{\partial f_2}{\partial x}(x,y) & \frac{\partial f_2}{\partial y}(x,y) \end{pmatrix}.$$

The linear Taylor expansion of $F(x,y)$ around the point $(a,b) \in \mathbb{R}^2$,

$$F(x,y) \approx F(a,b) + DF(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix},$$

is a second order approximation:

$$F(x,y) = F(a,b) + DF(a,b)\begin{pmatrix} x-a \\ y-b \end{pmatrix} + O\left(|x-a|^2\right) + O\left(|y-b|^2\right),$$

as $(x,y) \to (a,b)$.

5.3 Taylor series expansions

When discussing the approximation properties of Taylor polynomials in section 5.1, we assumed that the function $f : \mathbb{R} \to \mathbb{R}$ is differentiable a finite number of times. If the function $f(x)$ is infinitely many times differentiable, then we can define its Taylor series expansion as

$$T(x) = \sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} f^{(k)}(a). \tag{5.42}$$

Recall from (5.2) that $P_n(x) = \sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a)$, and therefore

$$T(x) = \lim_{n\to\infty} P_n(x). \tag{5.43}$$

In other words, the Taylor series expansion of the function $f(x)$ can also be viewed as the limit of the Taylor polynomials $P_n(x)$ corresponding to $f(x)$, as the degree n goes to infinity (at all points x where this limit exists).

Deciding when does the Taylor series $T(x)$ converge to the function $f(x)$, i.e., finding x such that $T(x) = f(x)$, is rather subtle.

Example: The function

$$f(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

is infinitely times many differentiable at all points, including at $x = 0$, and

$$f^{(k)}(0) = 0, \quad \forall\, k \geq 0.$$

Then, the Taylor series expansion of $f(x)$ around the point $a = 0$ is identically equal to zero, i.e.,

$$T(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} f^{(k)}(0) = 0, \quad \forall\, x \in \mathbb{R}.$$

Therefore, the Taylor expansion $T(x)$ is different than $f(x)$ at all the points $x \neq 0$. □

Note that the Taylor series $T(x)$ is a power series of the form (4.29), i.e.,

$$T(x) = \sum_{k=0}^{\infty} a_k (x-a)^k, \quad \text{with} \quad a_k = \frac{f^{(k)}(a)}{k!}, \quad \forall\, k \geq 0.$$

A practical way of computing the radius of convergence of a Taylor series, and therefore of finding the convergence domain of the series, is given in

Theorem 5.4. Before stating this theorem, we recall that, given a sequence $(x_n)_{n=1:\infty}$ of real numbers, $\liminf_{n\to\infty} x_n$ is defined as the smallest value l for which there exists a subsequence of $(x_n)_{n=1:\infty}$ converging to l. Note that,

$$\text{if } \lim_{n\to\infty} x_n \text{ exists, then } \liminf_{n\to\infty} x_n = \lim_{n\to\infty} x_n. \tag{5.44}$$

Also, if all the terms of the sequence $(x_n)_{n=1:\infty}$ are positive numbers, i.e., if $x_n > 0$ for all $n \geq 1$, then

$$\frac{1}{\limsup_{n\to\infty} x_n} = \liminf_{n\to\infty} \frac{1}{x_n}. \tag{5.45}$$

Theorem 5.4. *Let $f(x)$ be an infinitely many times differentiable function, and let $T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$ be the Taylor series expansion of $f(x)$ around the point a. The radius of convergence of $T(x)$ is given by*

$$R = \frac{1}{e} \liminf_{k\to\infty} \frac{k}{|f^{(k)}(a)|^{1/k}}. \tag{5.46}$$

If $\lim_{k\to\infty} \frac{k}{|f^{(k)}(a)|^{1/k}}$ exists, then

$$R = \frac{1}{e} \lim_{k\to\infty} \frac{k}{|f^{(k)}(a)|^{1/k}}. \tag{5.47}$$

Proof. By definition (4.33), the radius of convergence R of the Taylor series $T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$ is given by

$$R = \frac{1}{\limsup_{k\to\infty} \left| \frac{f^{(k)}(a)}{k!} \right|^{1/k}}.$$

Then, from (5.45), it follows that

$$R = \liminf_{k\to\infty} \left| \frac{k!}{f^{(k)}(a)} \right|^{1/k} \tag{5.48}$$

Using (4.37), we find that

$$\begin{aligned} \liminf_{k\to\infty} \left| \frac{k!}{f^{(k)}(a)} \right|^{1/k} &= \liminf_{k\to\infty} \left(\frac{k}{|f^{(k)}(a)|^{1/k}} \right) \cdot \left(\frac{(k!)^{1/k}}{k} \right) \\ &= \frac{1}{e} \liminf_{k\to\infty} \frac{k}{|f^{(k)}(a)|^{1/k}}. \end{aligned} \tag{5.49}$$

Then, (5.46) follows easily from (5.48) and (5.49), i.e.,

$$R = \frac{1}{e} \liminf_{k\to\infty} \frac{k}{|f^{(k)}(a)|^{1/k}}.$$

The result of (5.47) is obtained immediately from (5.46) using (5.44)[1]. □

The convergence and smoothness properties of the Taylor series expansions can be derived from Theorem 4.4 and Theorem 4.5:

Theorem 5.5. *Let $f(x)$ be an infinitely differentiable function, and let $T(x)$ be the Taylor series expansion of $f(x)$ around the point a. If R is the radius of convergence of $T(x)$, then $T(x)$ is convergent, and therefore well defined, on the interval $(a - R,\ a + R)$. Moreover, $T(x)$ is infinitely many times differentiable on the interval $(a - R,\ a + R)$.*

We note that Theorem 5.5 specifies the convergence domain of the Taylor series $T(x)$, but does not imply that $T(x) = f(x)$ on that domain. Sufficient conditions for the convergence of the Taylor series expansion $T(x)$ to the function $f(x)$ are given below.

Theorem 5.6. *Let $f(x)$ be an infinitely differentiable function, and let*

$$T(x) = \sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} f^{(k)}(a)$$

be the Taylor series expansion of $f(x)$ around the point a. Let R be the radius of convergence of $T(x)$, and let r such that $0 < r < R$. If

$$\lim_{n\to\infty} \frac{r^n}{n!} \sup_{z\in[a-r,a+r]} \left|f^{(n)}(z)\right| = 0, \tag{5.50}$$

then

$$T(x) = f(x), \quad \forall\ |x-a| \le r.$$

Proof. Since $r < R$, the Taylor series $T(x)$ is convergent for any x such that $|x - a| \le r$. Recall from (5.43) that $T(x) = \lim_{n\to\infty} P_n(x)$. Thus,

$$f(x) - T(x) = \lim_{n\to\infty} (f(x) - P_n(x)), \quad \forall\ |x-a| \le r. \tag{5.51}$$

Let $x \in [a - r, a + r]$ arbitrary. From the derivative form (5.3) for the Taylor residual, it follows that there exists a point c between a and x such that

$$f(x) - P_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c). \tag{5.52}$$

[1]We implicitly assumed that $f^{(k)}(a) \neq 0$, for $k \ge 0$. If this is not the case, the proof of Theorem 5.4 follows along the same lines, if we consider that $\frac{1}{|f^{(k)}(a)|} = \infty$, if $f^{(k)}(a) = 0$.

Note that $c \in [a - r, a + r]$, since $x \in [a - r, a + r]$. Then,

$$|f(x) - P_n(x)| = \frac{|x-a|^{n+1}}{(n+1)!} \left|f^{(n+1)}(c)\right| \le \frac{r^{n+1}}{(n+1)!} \sup_{z\in[a-r,a+r]} \left|f^{(n+1)}(z)\right|,$$

since $|x - a| \le r$. From (5.50), we obtain that

$$\lim_{n\to\infty} |f(x) - P_n(x)| = 0, \quad \forall\ x \in [a-r, a+r],$$

and therefore $T(x) = \lim_{n\to\infty} P_n(x) = f(x)$, for all $|x - a| \le r$. □

5.3.1 Examples of Taylor series expansions

The Taylor series expansions of the following functions are often used in practice:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots, \quad \forall\ x \in \mathbb{R}; \tag{5.53}$$

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots, \quad \forall\ x \in (-1, 1];$$

$$\ln(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \ldots, \quad \forall\ x \in [-1, 1); \tag{5.54}$$

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k = 1 - x + x^2 - x^3 + \ldots, \quad \forall\ x \in (-1, 1);$$

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \ldots, \quad \forall\ x \in (-1, 1). \tag{5.55}$$

All these Taylor series expansions are around 0; for example, (5.53) corresponds to (5.42) for $f(x) = e^x$ and $a = 0$.

Note that $(\ln(1-x))' = -\frac{1}{1-x}$. From Theorem 4.5, it follows that the Taylor series expansion of the function $\frac{1}{1-x}$ can be obtained by differentiating term by term the Taylor series expansion of the function $\ln(1 - x)$ and multiplying the result by (-1). This is, indeed, the case; see (5.54) and (5.55).

We conclude by analyzing the convergence of the Taylor series for e^x and $\ln(1 + x)$.

Example: The Taylor series expansion

$$T(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{5.56}$$

of the function $f(x) = e^x$ is convergent for all $x \in \mathbb{R}$.

Proof: We first check that the expansion (5.56) is correct. It is easy to see that $f^{(k)}(x) = e^x$, for all $k \geq 0$. Then $f^{(k)}(0) = 1$, for all $k \geq 0$, and the Taylor series expansion of $f(x)$ around the point $a = 0$ is

$$T(x) = \sum_{k=1}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=1}^{\infty} \frac{x^k}{k!}.$$

From (5.47) of Theorem 5.4, we obtain that the radius of convergence of the series $T(x)$ is

$$R = \frac{1}{e} \lim_{k\to\infty} \frac{k}{|f^{(k)}(0)|^{1/k}} = \lim_{k\to\infty} k = \infty.$$

We conclude that $T(x)$ is convergent for any $x \in \mathbb{R}$; cf. Theorem 4.4.

We use Theorem 5.6 to show that $f(x) = T(x)$ for any $x \in \mathbb{R}$.

Let $r > 0$ be a fixed arbitrary positive number. Note that

$$\sup_{x\in[-r,\, r]} |f^{(n)}(x)| = \sup_{-r\leq x\leq r} e^x = e^r.$$

Using (1.31), we conclude that

$$\lim_{n\to\infty} \frac{r^n}{n!} \sup_{x\in[-r,\, r]} |f^{(n)}(x)| = \lim_{n\to\infty} \frac{r^n}{n!} e^r = e^r \lim_{n\to\infty} \frac{r^n}{n!} = 0.$$

Thus, the convergence condition (5.50) is satisfied for any $r > 0$. From Theorem 5.6, it follows that $f(x) = T(x)$ for any real number x, i.e.,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad \forall\, x \in \mathbb{R}. \quad \square$$

Example: The Taylor series expansion

$$T(x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}, \tag{5.57}$$

of the function $\ln(1 + x)$ is convergent to $\ln(1 + x)$ for any $x \in [-0.5, 0.5]$.

Proof: We first check that the expansion (5.57) is correct. Let $f(x) = \ln(1 + x)$. It is easy to see, e.g., by induction, that

$$f^{(k)}(x) = \frac{(-1)^{k+1}\,(k-1)!}{(1+x)^k}, \quad \forall\, k \geq 1. \tag{5.58}$$

Then, $f^{(k)}(0) = (-1)^{k+1}(k-1)!$, for all $k \geq 1$. Since $f(0) = \ln(1) = 0$, it follows that the Taylor series of $f(x)$ around the point $a = 0$ is

$$T(x) = \sum_{k=1}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\,(k-1)!}{k!} x^k = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k,$$

which is the same as (5.57).

The series $T(x)$ can be written as a power series as follows:

$$T(x) = \sum_{k=1}^{\infty} a_k x^k, \quad \text{with} \quad a_k = \frac{(-1)^{k+1}}{k}, \quad \forall\, k \geq 1.$$

Recall from (1.29) that $\lim_{k\to\infty} k^{1/k} = 1$. Therefore,

$$\lim_{k\to\infty} |a_k|^{1/k} = \lim_{k\to\infty} \left(\frac{1}{k}\right)^{1/k} = \lim_{k\to\infty} \frac{1}{k^{1/k}} = 1,$$

From Lemma 4.9, it follows that the radius of convergence of $T(x)$ is

$$R = \frac{1}{\lim_{k\to\infty} |a_k|^{1/k}} = 1.$$

We conclude from Theorem 4.4 that $T(x)$ is convergent if $|x| < 1$, and that $T(x)$ is not convergent if $|x| > 1$.

We now use Theorem 5.6 to find points x where $T(x) = f(x)$.

Let $0 < r < 1$. From (5.58), we find that

$$\sup_{x\in[-r,\, r]} |f^{(n)}(x)| = \sup_{-r\leq x\leq r} \frac{(n-1)!}{(1+x)^n} = \frac{(n-1)!}{(1-r)^n}.$$

Therefore,

$$\lim_{n\to\infty} \frac{r^n}{n!} \sup_{x\in[-r,\, r]} |f^{(n)}(x)| = \lim_{n\to\infty} \frac{r^n}{n!} \frac{(n-1)!}{(1-r)^n} = \lim_{n\to\infty} \frac{1}{n} \left(\frac{r}{1-r}\right)^n.$$

If $r \leq \frac{1}{2}$, then $\frac{r}{1-r} \leq 1$ and we obtain that

$$\lim_{n\to\infty} \frac{r^n}{n!} \sup_{x\in[-r,\, r]} |f^{(n)}(x)| = 0.$$

The convergence condition (5.50) is satisfied in this case. From Theorem 5.6, it follows that $f(x) = T(x)$, for any x with $|x| \leq \frac{1}{2}$, i.e.,

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}, \quad \forall\, |x| \leq \frac{1}{2}. \tag{5.59}$$

The result of Theorem 5.6 is not strong enough to prove that (5.59) also holds[2] if $\frac{1}{2} < |x| < 1$. To do so, the Taylor approximation error must be estimated using the integral form (5.4); more details are provided in an exercise at the end of the chapter. □

FINANCIAL APPLICATIONS

Connections between the Greeks using Taylor's formula.

Approximation of the Black–Scholes formula for at–the–money options.

Connections between the duration and convexity of a bond.

5.4 Connections between the Greeks using Taylor's formula

Let $V(S,t)$ be the value at time t of a European plain vanilla option on an underlying asset with spot price S. Recall that $V(S,t)$ is an infinitely many times differentiable function in both variables S and t.

To write the quadratic Taylor expansion of $V(S,t)$ around the point (S,t) we use (5.39) for $a = S$, $b = t$, $x = S + dS$, and $y = t + dt$, and obtain that

$$\begin{aligned} V(S+dS,t+dt) &= V(S,t) + dS\frac{\partial V}{\partial S} + dt\frac{\partial V}{\partial t} \\ &+ \frac{(dS)^2}{2}\frac{\partial^2 V}{\partial S^2} + \frac{(dt)^2}{2}\frac{\partial^2 V}{\partial t^2} + (dS)(dt)\frac{\partial^2 V}{\partial S \partial t} \\ &+ O((dS)^3) + O((dt)^3). \end{aligned} \tag{5.60}$$

For the discretized evolution model of the price of the underlying asset to converge to a lognormal model, the following condition must be satisfied:

$$(dS)^2 \approx \sigma^2 S^2 dt.$$

Let $dV = V(S+dS,t+dt) - V(S,t)$. If we ignore all the terms of order larger than dt from (5.60), we find that

$$dV \approx \frac{\partial V}{\partial S}\,dS + \frac{\partial V}{\partial t}\,dt + \frac{(dS)^2}{2}\frac{\partial^2 V}{\partial S^2}$$

[2]We also note that $f(-1)$ and $T(-1)$ are not defined, and that $T(1) = f(1)$:

$$T(1) = \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k} = \ln(2) = f(1).$$

$$\approx \frac{\partial V}{\partial S}\,dS + \frac{\partial V}{\partial t}\,dt + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2}\,dt. \tag{5.61}$$

Recall that

$$\Delta = \frac{\partial V}{\partial S}, \quad \Gamma = \frac{\partial^2 V}{\partial S^2}, \quad \text{and} \quad \Theta = \frac{\partial V}{\partial t}.$$

Then, the approximate formula (5.61) can be written as

$$dV \approx \Delta\,dS + \Theta\,dt + \frac{\sigma^2 S^2}{2}\Gamma\,dt,$$

or, equivalently,

$$dV - \Delta\,dS \approx \Theta\,dt + \frac{\sigma^2 S^2}{2}\Gamma\,dt. \tag{5.62}$$

For simplicity, assume that the underlying asset does not pay dividends.

To hedge a long position in the derivative security, a short position in Δ units of the underlying asset must be taken. If Π is the value of the resulting portfolio, then $\Pi = V - \Delta S$. Note that if the value of the underlying asset changes from S to $S+dS$, then the value of the portfolio changes as follows[3]:

$$\begin{aligned} d\Pi &= \Pi(S+dS) - \Pi(S) \\ &= (V(S+dS) - \Delta(S+dS)) - (V(S) - \Delta S) \\ &= V(S+dS) - V(S) - \Delta dS \\ &= dV - \Delta dS. \end{aligned}$$

Then, (5.62) becomes

$$d\Pi \approx \Theta\,dt + \frac{\sigma^2 S^2}{2}\Gamma\,dt. \tag{5.63}$$

The reason for shorting Δ units of the underlying asset in the portfolio is to make the value of the portfolio insensitive to small changes in the price of the underlying asset. Since

$$\frac{\partial \Pi}{\partial S} = \frac{\partial V}{\partial S} - \Delta = 0,$$

it follows that the value of the portfolio over the small time period dt corresponding to the change in the price of the underlying asset from S to $S+dS$

[3]If the underlying asset paid dividends continuously at the rate q, then

$$d\Pi = dV - \Delta dS - q\Delta S dt.$$

is fully deterministic. To avoid arbitrage opportunities, the value of this portfolio must grow at the same rate as the risk free interest rate r, i.e.,

$$d\Pi \; = \; r\Pi \, dt. \qquad (5.64)$$

From (5.63) and (5.64), we conclude that

$$r\Pi \; \approx \; \Theta \; + \; \frac{\sigma^2 S^2}{2}\Gamma. \qquad (5.65)$$

We note that the approximate formula (5.65) is actually an equality for plain vanilla European options; cf. (6.53) for $q = 0$. This equality can be derived using the Black–Scholes PDE; see section 6.4.2 for details.

5.5 Approximation of the Black–Scholes formula for at–the–money options

For at-the-money options, approximations to the Black–Scholes formula that do not require estimating the cumulative density function of the standard normal distributions can be derived. These formulas are easy to compute and, in many cases, very accurate.

In section 5.5.1, we present several such approximations formulas, including approximations for the Greeks and for the implied volatility. These formulas are proved in section 5.5.2. In section 5.5.3, we derive a theoretical bound for the precision of the approximation of the price for ATM options on non–dividend–paying assets for the case of zero interest rates. We also investigate the relative approximation error of the formulas from section 5.5.1 for two particular examples.

5.5.1 Several ATM approximations formulas

The most commonly used approximation formulas are for at-the-money options on assets paying no dividends, with zero risk-free rates:

$$\boxed{r = 0;\; q = 0: \quad P \;=\; C \;\approx\; \sigma S\sqrt{\frac{T}{2\pi}} \;\approx\; 0.4\,\sigma S\sqrt{T}}$$

For the general case when r and q are nonzero, the following approximation formulas can be used for pricing ATM plain vanilla European options:

$$r \neq 0;\; q \neq 0: \quad C \;\approx\; \sigma S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{(r+q)T}{2}\right) \;+\; \frac{(r-q)T}{2}S$$

$$r \neq 0;\; q \neq 0: \quad P \;\approx\; \sigma S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{(r+q)T}{2}\right) \;-\; \frac{(r-q)T}{2}S$$

Approximation formulas can also be obtained for the implied volatility of ATM options:

$$r \;=\; q \;=\; 0: \quad \sigma_{imp} \;\approx\; \frac{C\sqrt{2\pi}}{S\sqrt{T}} \;\approx\; \frac{2.5\;C}{S\sqrt{T}}$$

$$r \neq 0;\; q \neq 0: \quad \sigma_{imp} \;\approx\; \frac{\sqrt{2\pi}}{S\sqrt{T}}\,\frac{C - \frac{(r-q)T}{2}S}{1-\frac{(r+q)T}{2}} \;\approx\; \frac{\sqrt{2\pi}}{S\sqrt{T}}\,\frac{P + \frac{(r-q)T}{2}S}{1-\frac{(r+q)T}{2}}$$

If $r = q = 0$, the following approximation formulas can be used for the Greeks of ATM options (similar, but more complicated formulas exist for $r, q \neq 0$):

$$\Delta(C) \;\approx\; \frac{1}{2} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}} \;\approx\; \frac{1}{2} + 0.2\sigma\sqrt{T}$$

$$\Delta(P) \;\approx\; -\frac{1}{2} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}} \;\approx\; -\frac{1}{2} + 0.2\sigma\sqrt{T}$$

$$\Gamma(C) \;=\; \Gamma(P) \;\approx\; \frac{1}{\sigma S\sqrt{2\pi T}}\left(1 - \frac{\sigma^2 T}{8}\right) \;\approx\; \frac{0.4}{\sigma S\sqrt{T}}\left(1 - \frac{\sigma^2 T}{8}\right)$$

$$\text{vega}(C) \;=\; \text{vega}(P) \;\approx\; S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{\sigma^2 T}{8}\right) \;\approx\; 0.4S\sqrt{T}\left(1 - \frac{\sigma^2 T}{8}\right)$$

$$\Theta(C) \;=\; \Theta(P) \;\approx\; -\frac{\sigma S}{2\sqrt{2\pi T}}\left(1 - \frac{\sigma^2 T}{8}\right) \;\approx\; -\frac{0.2\sigma S}{\sqrt{T}}\left(1 - \frac{\sigma^2 T}{8}\right)$$

$$\rho(C) \;\approx\; ST\left(\frac{1}{2} - \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}}\right) \;\approx\; ST\left(\frac{1}{2} - 0.2\sigma\sqrt{T}\right)$$

$$\rho(P) \;\approx\; -ST\left(\frac{1}{2} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}}\right) \;\approx\; -ST\left(\frac{1}{2} + 0.2\sigma\sqrt{T}\right)$$

5.5.2 Deriving the ATM approximations formulas

We begin by proving that, if $r = 0$ and $q = 0$, the prices of the ATM call and ATM put are equal, and can be approximated as

$$P \;=\; C \;\approx\; \sigma S\sqrt{\frac{T}{2\pi}}, \qquad (5.66)$$

or, in a simpler to compute but less precise way, as

$$P \;=\; C \;\approx\; 0.4\,\sigma S\sqrt{T}, \qquad (5.67)$$

since $\frac{1}{\sqrt{2\pi}} \;=\; 0.39894228 \;\approx\; 0.4.$

Since $r = q = 0$ and $K = S$, it is easy to see from the Put–Call parity (1.47) that the values of the put option and call options are the same:

$$P = Ke^{-rT} - S + C = S - S + C = C. \tag{5.68}$$

For an ATM call and assuming zero interest rates and a non-dividend-paying asset, i.e., for $K = S$, $r = 0$, and $q = 0$, the Black–Scholes formula (3.53) corresponding to $t = 0$ can be written as

$$C = SN\left(\frac{\sigma\sqrt{T}}{2}\right) - SN\left(-\frac{\sigma\sqrt{T}}{2}\right). \tag{5.69}$$

From (3.43), we find that $N\left(-\frac{\sigma\sqrt{T}}{2}\right) = 1 - N\left(\frac{\sigma\sqrt{T}}{2}\right)$ and (5.69) becomes

$$C = 2S\left(N\left(\frac{\sigma\sqrt{T}}{2}\right) - \frac{1}{2}\right). \tag{5.70}$$

We approximate the term $N\left(\frac{\sigma\sqrt{T}}{2}\right)$ using a Taylor approximation around 0. From (5.7) for $f(x) = N(x)$, $a = 0$, and $n = 2$, we obtain that

$$N(x) = N(0) + xN'(0) + \frac{x^2}{2}N''(0) + O(x^3), \tag{5.71}$$

as $x \to 0$. Recall that

$$N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}}\, dy.$$

It is easy to see that $N(0) = \frac{1}{2}$; cf. (3.43) for $a = 0$. From Lemma 2.3, we obtain that $N'(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$, and therefore $N''(t) = -\frac{1}{\sqrt{2\pi}}te^{-\frac{t^2}{2}}$. Thus, $N(0) = \frac{1}{2}$, $N'(0) = \frac{1}{\sqrt{2\pi}}$, $N''(0) = 0$, and the Taylor expansion (5.71) of $N(x)$ around the point 0 becomes[4]

$$N(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} + O(x^3) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}}. \tag{5.72}$$

[4]We note that the Taylor expansion (5.72) can be more accurately written as

$$N(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} - \frac{x^3}{6}\frac{1}{\sqrt{2\pi}} + O(x^5), \quad \text{as } x \to 0,$$

since $N'''(0) = -\frac{1}{\sqrt{2\pi}}$, $N^{(4)}(0) = 0$, and $N^{(5)}(0) = \frac{3}{\sqrt{2\pi}} \neq 0$. This level of precision is not needed here.

Thus,

$$N\left(\frac{\sigma\sqrt{T}}{2}\right) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\frac{\sigma\sqrt{T}}{2}, \tag{5.73}$$

and, by substituting (5.73) into (5.70), we find that

$$C = 2S\left(N\left(\frac{\sigma\sqrt{T}}{2}\right) - \frac{1}{2}\right) \approx \sigma S\sqrt{\frac{T}{2\pi}},$$

which is the same as (5.66).

By solving (5.66) for σ, we obtain the following estimate for the implied volatility of an ATM call option:

$$\sigma_{imp} \approx \sqrt{2\pi}\frac{C}{S\sqrt{T}} \approx \frac{2.5\, C}{S\sqrt{T}}; \tag{5.74}$$

see sections 3.6.2 and 8.6 for more details on implied volatility.

Formula (5.66) is a good approximation for the Black–Scholes value of ATM options only if the Taylor approximation (5.72) is accurate. In section 5.5.3, we show that, for options with maturities less than one year and volatility slightly smaller than 50%, the approximation given by (5.66) is within one percent of the Black–Scholes option price; see Theorem 5.7 for the precise result.

The following approximations of the Black–Scholes formulas for at–the–money options if $r \neq 0$ and $q \neq 0$, i.e.,

$$C \approx \sigma S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{(r+q)T}{2}\right) + \frac{(r-q)T}{2}S; \tag{5.75}$$

$$P \approx \sigma S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{(r+q)T}{2}\right) - \frac{(r-q)T}{2}S, \tag{5.76}$$

can be obtained similarly. We only derive (5.75) here. For an ATM call option, the Black–Scholes formula (3.53) corresponding to $t = 0$ becomes

$$C = Se^{-qT}N\left(\frac{(r-q)\sqrt{T}}{\sigma} + \frac{\sigma\sqrt{T}}{2}\right) - Se^{-rT}N\left(\frac{(r-q)\sqrt{T}}{\sigma} - \frac{\sigma\sqrt{T}}{2}\right).$$

From the Taylor expansion (5.72) for $N(x)$, i.e., $N(x) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}}$, we find that

$$C = Se^{-qT}\left(\frac{1}{2} + \frac{(r-q)\sqrt{T}}{\sigma\sqrt{2\pi}} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}}\right)$$

$$
\begin{aligned}
& - Se^{-rT}\left(\frac{1}{2} + \frac{(r-q)\sqrt{T}}{\sigma\sqrt{2\pi}} - \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}}\right) \\
= \; & S\left(\frac{1}{2} + \frac{(r-q)\sqrt{T}}{\sigma\sqrt{2\pi}}\right)\left(e^{-qT} - e^{-rT}\right) + \sigma S\sqrt{\frac{T}{2\pi}}\,\frac{e^{-qT}+e^{-rT}}{2}.
\end{aligned}
$$

Using the quadratic Taylor approximations (5.20) for e^{-qT} and e^{-rT}, i.e., $e^{-qT} \approx 1 - qT + \frac{q^2T^2}{2}$ and $e^{-rT} \approx 1 - rT + \frac{r^2T^2}{2}$, we find that

$$
\begin{aligned}
e^{-qT} - e^{-rT} &= (r-q)T - \frac{(r^2-q^2)T^2}{2}; \\
\frac{e^{-qT}+e^{-rT}}{2} &= 1 - \frac{(r+q)T}{2} + \frac{(r^2+q^2)T^2}{4},
\end{aligned}
$$

and obtain that

$$
\begin{aligned}
C \approx \; & S\frac{(r-q)T}{2} + S\frac{(r-q)^2T\sqrt{T}}{\sigma\sqrt{2\pi}} - S\frac{(r^2-q^2)T^2}{4} \\
& - S\frac{(r-q)(r^2-q^2)T^2\sqrt{T}}{2\sigma\sqrt{2\pi}} + \sigma S\sqrt{\frac{T}{2\pi}}\left(1 - \frac{(r+q)T}{2} + \frac{(r^2+q^2)T^2}{4}\right).
\end{aligned}
$$

When ignoring the terms involving r^2 and q^2, which are, in general, much smaller than the other terms, the approximation formula (5.75) is obtained.

Solving for σ in (5.75), the following estimate for the implied volatility of ATM calls when $r \neq 0$ and $q \neq 0$ is found:

$$
\sigma_{imp} \approx \frac{\sqrt{2\pi}}{S\sqrt{T}}\,\frac{C - \frac{(r-q)T}{2}S}{1 - \frac{(r+q)T}{2}}. \tag{5.77}
$$

To obtain the ATM approximations of the Greeks of ATM plain vanilla European options, we use the closed formulas (3.66–3.75). We only present the formulas for the case $r = q = 0$. For the general case $r \neq 0$ and $q \neq 0$, the approximation formulas for the Greeks are obtained similarly.

If $K = S$ and $r = q = 0$, and for $t = 0$, the formulas (3.55) and (3.56) become

$$
d_1 = \frac{\sigma\sqrt{T}}{2} \quad \text{and} \quad d_2 = -d_1. \tag{5.78}
$$

Recall that

$$
\frac{1}{\sqrt{2\pi}} = 0.39894228 \approx 0.4. \tag{5.79}
$$

From (5.72), i.e., $N(x) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}}$, we find that

$$
N(d_1) \approx \frac{1}{2} + \frac{d_1}{\sqrt{2\pi}} = \frac{1}{2} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}} \approx \frac{1}{2} + 0.2\sigma\sqrt{T}; \tag{5.80}
$$

$$
N(-d_1) \approx \frac{1}{2} - \frac{d_1}{\sqrt{2\pi}} = \frac{1}{2} - \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}} \approx \frac{1}{2} - 0.2\sigma\sqrt{T}. \tag{5.81}
$$

From (5.78) and the first order Taylor approximation (5.15) for the function e^x, we obtain that

$$
e^{-\frac{d_1^2}{2}} \approx 1 - \frac{d_1^2}{2} = 1 - \frac{\sigma^2T}{8}. \tag{5.82}
$$

Using (5.79–5.82), we conclude from (3.66–3.75) for $t = 0$, $K = S$ and $r = q = 0$ that

$$
\begin{aligned}
\Delta(C) &= N(d_1) \approx \frac{1}{2} + 0.2\sigma\sqrt{T}; \\
\Delta(P) &= -N(-d_1) \approx -\frac{1}{2} + 0.2\sigma\sqrt{T}; \\
\Gamma(C) = \Gamma(P) &= \frac{1}{S\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}} \approx \frac{0.4}{S\sigma\sqrt{T}}\left(1 - \frac{\sigma^2T}{8}\right); \\
\mathrm{vega}(C) = \mathrm{vega}(P) &= S\sqrt{T}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}} \approx 0.4S\sqrt{T}\left(1 - \frac{\sigma^2T}{8}\right); \\
\Theta(C) = \Theta(P) &= -\frac{\sigma S}{2\sqrt{T}}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}} \approx -\frac{0.2\sigma S}{\sqrt{T}}\left(1 - \frac{\sigma^2T}{8}\right); \\
\rho(C) &= ST\,N(-d_1) \approx ST\left(\frac{1}{2} - 0.2\sigma\sqrt{T}\right); \\
\rho(P) &= -ST\,N(d_1) \approx -ST\left(\frac{1}{2} + 0.2\sigma\sqrt{T}\right).
\end{aligned}
$$

5.5.3 The precision of the ATM approximation of the Black–Scholes formula

A theoretical estimate for when the approximate value given by formula (5.66) for the price of the option of an ATM call on a non–dividend–paying asset is derived in Theorem 5.7. The following result will be needed in the proof of Theorem 5.7:

Lemma 5.1. *Let $N(x)$ be the cumulative density function of the standard normal variable. Then,*

$$N\left(\frac{\sigma\sqrt{T}}{2}\right) = \frac{1}{2} + \frac{\sigma\sqrt{T}}{2\sqrt{2\pi}} - \frac{1}{\sqrt{2\pi}}\int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, e^{-\frac{t^2}{2}}\, dt. \quad (5.83)$$

Proof. For $n = 1$, formula (5.4) for the integral form of the Taylor approximation becomes

$$f(x) - f(a) - (x-a)f'(a) = \int_a^x (x-t)\, f''(t)\, dt, \quad (5.84)$$

since $P_1(x) = f(a) + (x-a)f'(a)$. By writing (5.84) for $f(x) = N(x)$ and $a = 0$, we obtain

$$N(x) - N(0) - xN'(0) = \int_0^x (x-t)\, N''(t)\, dt.$$

Recall that $N(0) = \frac{1}{2}$, $N'(0) = \frac{1}{\sqrt{2\pi}}$, and $N''(t) = -\frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}}$. Then,

$$N(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} - \frac{1}{\sqrt{2\pi}}\int_0^x (x-t)\, t\, e^{-\frac{t^2}{2}}\, dt. \quad (5.85)$$

Formula (5.83) follows from (5.85), by substituting $x = \frac{\sigma\sqrt{T}}{2}$. □

Theorem 5.7. *Consider an at–the–money call option on a non–dividend paying asset, and assume zero interest rates. Let $C_{BS,r=q=0}$ be the Black–Scholes value of the call, and let $C_{approx,r=q=0}$ be the approximate value of the call option given by (5.66). The relative error given by the approximation formula (5.66) for ATM calls on non–dividend–paying assets is less than 1%, i.e.,*

$$\frac{|C_{approx,r=q=0} - C_{BS,r=q=0}|}{C_{BS,r=q=0}} \leq 0.01, \quad (5.86)$$

provided that the total variance $\sigma^2 T$ is bounded as follows:

$$\sigma^2 T \leq \frac{24}{101} \approx 0.2376.$$

Proof. Recall from (5.66) that

$$C_{approx,r=q=0} = \sigma S \frac{\sqrt{T}}{\sqrt{2\pi}}. \quad (5.87)$$

From (5.70), and using (5.83), we find that

$$\begin{aligned} C_{BS,r=q=0} &= 2S\left(N\left(\frac{\sigma\sqrt{T}}{2}\right) - \frac{1}{2}\right) \\ &= S\frac{\sigma\sqrt{T}}{\sqrt{2\pi}} - \frac{2S}{\sqrt{2\pi}}\int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, e^{-\frac{t^2}{2}}\, dt \\ &= C_{approx,r=q=0} - \frac{2S}{\sqrt{2\pi}}\int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, e^{-\frac{t^2}{2}}\, dt. \quad (5.88) \end{aligned}$$

From (5.88), we obtain that $C_{BS,r=q=0} < C_{approx,r=q=0}$. Then, condition (5.86) is equivalent to

$$\frac{|C_{approx,r=q=0} - C_{BS,r=q=0}|}{C_{BS,r=q=0}} = \frac{C_{approx,r=q=0} - C_{BS,r=q=0}}{C_{BS,r=q=0}} \leq 0.01,$$

and therefore to

$$C_{BS,r=q=0} \geq \frac{1}{1.01} C_{approx,r=q=0}. \quad (5.89)$$

Using again (5.88) and (5.87), the inequality (5.89) can be written as

$$\frac{2S}{\sqrt{2\pi}}\int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, e^{-\frac{t^2}{2}}\, dt \leq \frac{0.01}{1.01} C_{approx,r=q=0} = \frac{1}{101}\sigma\sqrt{T}\frac{S}{\sqrt{2\pi}}. \quad (5.90)$$

Since $e^{-\frac{t^2}{2}} < 1$, it is easy to see that

$$\int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, e^{-\frac{t^2}{2}}\, dt < \int_0^{\frac{\sigma\sqrt{T}}{2}} \left(\frac{\sigma\sqrt{T}}{2} - t\right) t\, dt = \frac{(\sigma\sqrt{T})^3}{48}.$$

Therefore, the inequality (5.90) is satisfied provided that

$$\frac{(\sigma\sqrt{T})^3}{48} \leq \frac{1}{101}\frac{\sigma\sqrt{T}}{2}.$$

This is the same as $\sigma^2 T \leq \frac{24}{101}$, which is what we wanted to prove. □

We conclude by investigating how good the approximations from section 5.5.1 are, if $r = 0$ and $q = 0$, and if $r \neq 0$ and $q \neq 0$, for two different ATM options.

Example: Estimate the relative errors given by the approximation formulas from section 5.5.1 for the value, the Greeks, and the implied volatility of

a six months at–the–money call with strike 40 on a non–dividend–paying underlying asset with volatility 25%, assuming zero interest rates.

Answer: For $S = K = 40$, $T = 0.5$, $\sigma = 0.25$, and $r = q = 0$, the Black-Scholes value of the call is $C_{BS,r=q=0} = 2.817284$, while the approximate value of the call computed using formula (5.66) is $C_{approx,r=q=0} = 2.820948$. The relative error of the approximation is

$$\frac{|C_{BS,r=q=0} - C_{approx,r=q=0}|}{C_{BS,r=q=0}} = 0.0013. \tag{5.91}$$

In other words, the approximate value is within 0.13% of the Black–Scholes value, which is an excellent approximation.

Computing the Greeks of the call option using the formulas from section 3.6, as well as the approximate values for the Greeks using the formulas from section 5.5.1 we obtain

$$\begin{array}{rclrcl} \Delta & = & 0.53521605; & \Delta\_\text{approx} & = & 0.53526185; \\ \Gamma & = & 0.05619900; & \Gamma\_\text{approx} & = & 0.05619857; \\ \text{vega} & = & 11.23980034; & \text{vega\_approx} & = & 11.23971436; \\ \Theta & = & -2.80995008; & \Theta\_\text{approx} & = & -2.80992859; \\ \rho & = & 9.29567892; & \rho\_\text{approx} & = & 9.29476302. \end{array}$$

Thus, the relative approximation errors for the Greeks of the call option are

$$\begin{aligned} \frac{|\Delta - \Delta\_\text{approx}|}{\Delta} &= 8.5564 \cdot 10^{-5}; \\ \frac{|\Gamma - \Gamma\_\text{approx}|}{\Gamma} &= 7.6493 \cdot 10^{-6}; \\ \frac{|\text{vega} - \text{vega\_approx}|}{\text{vega}} &= 7.6493 \cdot 10^{-6}; \\ \frac{|\Theta - \Theta\_\text{approx}|}{|\Theta|} &= 7.6493 \cdot 10^{-6}; \\ \frac{|\rho - \rho\_\text{approx}|}{\rho} &= 9.8530 \cdot 10^{-5}. \end{aligned}$$

We conclude that the approximations for the Greeks of the ATM options are very accurate[5], for this particular case.

[5]The relative approximation errors for Γ, vega, and Θ are supposed to be identical, since

$$\frac{|\Gamma - \Gamma\_\text{approx}|}{\Gamma} = \frac{\left|e^{-\frac{d_1^2}{2}} - \left(1 - \frac{\sigma^2 T}{8}\right)\right|}{e^{-\frac{d_1^2}{2}}} = \frac{|\text{vega} - \text{vega\_approx}|}{\text{vega}}.$$

We conclude by analyzing the precision of the approximation (5.74) for the implied volatility, i.e.,

$$\sigma_{imp,approx} = \sqrt{2\pi}\frac{C}{S\sqrt{T}}. \tag{5.92}$$

Using (5.92) with $C = C_{BS,r=q=0} = 2.817284$, we find that $\sigma_{imp,approx} = 0.249675$. Since the value $C_{BS,r=q=0}$ was obtained from the Black–Scholes formula using the value $\sigma = 0.25$ for the volatility, it follows that $\sigma = 0.25$ is the exact value of the implied volatility. The relative approximation error is

$$\frac{|\sigma - \sigma_{imp,approx}|}{\sigma} = 0.0013.$$

Thus, the implied volatility approximate value is within 0.13% of the volatility used to price the call option. □

Example: Estimate the relative errors given by the approximation formulas (5.75) and (5.77) for the value and the implied volatility of a six months at–the–money call with strike 40 on an underlying asset with volatility 25% paying dividends continously at rate 3%, is the risk–free interest rates are constant at 6%.

Answer: For $S = K = 40$, $T = 0.5$, $\sigma = 0.25$, $q = 0.03$, and $r = 0.06$, the value of the call given by the Black-Scholes formula is $C_{BS,r=0.06,q=0.03} = 3.057889$, while the approximate value computed using formula (5.75) is $C_{approx,r=0.06,q=0.03} = 3.057477$. The relative error of the approximation is

$$\frac{|C_{BS,r=0.06,q=0.03} - C_{approx,r=0.06,q=0.03}|}{C_{BS,r=0.06,q=0.03}} = 0.00013.$$

The approximation formula is within 0.01% of the Black–Scholes price for this case, and therefore extremely accurate.

We recall the approximation (5.77) for the implied volatility, i.e.,

$$\sigma_{imp,approx} = \frac{\sqrt{2\pi}}{S\sqrt{T}} \frac{C - \frac{(r-q)T}{2}S}{1 - \frac{(r+q)T}{2}}. \tag{5.93}$$

Using (5.93) with $C = C_{BS,r=0.06,q=0.03} = 3.057889$, the Black-Scholes value of the call computed previously, we find that $\sigma_{imp,approx} = 0.250037$. Since the exact value of the implied volatility is $\sigma = 0.25$, the relative approximation error for the implied volatility is

$$\frac{|\sigma - \sigma_{imp,approx}|}{\sigma} = 0.00015.$$

Thus, the implied volatility approximate value is within 0.015% of the volatility used to price the call option, which is remarkably good accuracy. □

5.6 Connections between duration and convexity

Recall from section 2.7 that bond duration measures the change in the price of a bond with respect to changes in the yield curve, while bond convexity measures the change of the duration of a bond with respect to changes in the yield curve, i.e.,

$$D = -\frac{1}{B}\frac{\partial B}{\partial y} \quad \text{and} \quad C = \frac{1}{B}\frac{\partial^2 B}{\partial y^2}. \tag{5.94}$$

Also, recall that the value B of a bond with yield y paying cash–flows c_i and time t_i, $i = 1:n$, is $B = \sum_{i=1}^n c_i e^{-yt_i}$. To emphasize that the value of the bond is a function of its yield, we denote B by $B(y)$, i.e.,

$$B(y) = \sum_{i=1}^n c_i e^{-yt_i}. \tag{5.95}$$

If the yield of the bond changes from y to $y+\Delta y$ over a small time interval Δt, then the new price of the bond can be computed as the sum of the discounted present values of all the cash flows computed using the new yield $y+\Delta y$. Since the cash flow c_i will no longer be received at time t_i, but at time $t_i - \Delta t$, we find that[6]

$$B(y+\Delta y) = \sum_{i=1}^n c_i e^{-(y+\Delta y)(t_i - \Delta t)}.$$

An alternative to discounting the future cash flows to compute the new price $B(y+\Delta y)$ of the bond is given below.

Lemma 5.2. *Let D and C be the duration and convexity of a bond with yield y and value $B = B(y)$. Then,*

$$\frac{\Delta B}{B} \approx -D\Delta y + \frac{1}{2}C(\Delta y)^2, \tag{5.96}$$

where $\Delta B = B(y+\Delta y) - B(y)$.

Note that using formula (5.96) does not require specific knowledge of the cash flows of the bond in order to compute an approximate value for the return $\frac{\Delta B}{B}$ of a long bond position.

Proof. We use the quadratic Taylor approximation (5.13), i.e.,

$$f(x) \approx f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2}f''(a)$$

[6]We implicitly assumed that $t_1 > 0$. Then, Δt can be chosen small enough such that $t_1 - \Delta t > 0$.

for the function $f(x) = B(y)$, and for the points $x = y + \Delta y$ and $a = y$:

$$B(y+\Delta y) \approx B(y) + \Delta y\, B'(y) + \frac{(\Delta y)^2}{2} B''(y). \tag{5.97}$$

Let $\Delta B = B(y+\Delta y) - B(y)$. Using the partial derivatives notation[7] for the derivatives of $B(y)$ with respect to y, i.e., $B'(y) = \frac{\partial B}{\partial y}$ and $B''(y) = \frac{\partial^2 B}{\partial y^2}$, the Taylor expansion (5.97) can be written as

$$\Delta B \approx \Delta y \frac{\partial B}{\partial y} + \frac{(\Delta y)^2}{2}\frac{\partial^2 B}{\partial y^2}.$$

From (5.94), it is easy to see that $\frac{\partial B}{\partial y} = -DB$ and $\frac{\partial^2 B}{\partial y^2} = CB$. Therefore,

$$\Delta B \approx -D\,B\,\Delta y + C\,B\,\frac{(\Delta y)^2}{2}, \tag{5.98}$$

and formula (5.96) is obtained by dividing both sides of (5.98) by B. □

Formula (5.96) can be used to justify a fact well known in bond trading: of two bonds with equal duration, the bond with higher convexity provides a higher return for small changes in the yield curve.

To see this formally, let B_i, D_i, and C_i denote the price, duration, and convexity of bond i, for $i = 1,2$. We know that $D_1 = D_2 = D$. Assume that $C_1 \geq C_2$. If the yield of both bonds changes by a small amount denoted by Δy, then the return of a long position in either one of the bonds is

$$\frac{B_i(y+\Delta y) - B_i(y)}{B_i(y)} = \frac{\Delta B_i}{B_i}, \quad i = 1,2.$$

From (5.96), we find that

$$\frac{\Delta B_1}{B_1} \approx -D_1\Delta y + \frac{1}{2}C_1(\Delta y)^2 = -D\Delta y + \frac{1}{2}C_1(\Delta y)^2;$$
$$\frac{\Delta B_2}{B_2} \approx -D_2\Delta y + \frac{1}{2}C_2(\Delta y)^2 = -D\Delta y + \frac{1}{2}C_2(\Delta y)^2.$$

Since $C_1 \geq C_2$, and since $(\Delta y)^2 \geq 0$, we conclude that, regardless of whether the yield goes up or down,

$$\frac{\Delta B_1}{B_1} \geq \frac{\Delta B_2}{B_2}.$$

In other words, the bond with higher convexity provides a higher return, and is the better investment, all other things being considered equal.

[7]One reason for using the partial derivatives notation is that the price of the bond B can be regarded not just as a function of the yield y, but also of, e.g., the cash flows c_i, $i = 1:n$; cf. (5.95).

5.7 References

Taylor's formulas for functions of one and two variable are covered in Protter and Morrey [20]. Some of the ATM approximation formulas from section 5.5 were introduced by Brenner and Subrahmanyam [3].

5.8 Exercises

1. Show that the cubic Taylor approximation of $\sqrt{1+x}$ around 0 is

$$\sqrt{1+x} \approx 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16}.$$

2. Use the Taylor series expansion of the function e^x to find the value of $e^{0.25}$ with six decimal digits accuracy.

3. Find the Taylor series expansion of the functions

$$\ln(1-x^2) \quad \text{and} \quad \frac{1}{1-x^2}$$

around the point 0, using the Taylor series expansions (5.54) and (5.55) of $\ln(1-x)$ and $\frac{1}{1-x}$.

4. Let

$$T(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\, x^k}{k}$$

be the Taylor series expansion of $f(x) = \ln(1+x)$. In section 5.3.1, we showed that $T(x) = f(x)$ if $|x| \le \frac{1}{2}$. In this exercise, we show that $T(x) = f(x)$ for all x such that $|x| < 1$.

Let $P_n(x)$ be the Taylor polynomial of degree n corresponding to $f(x)$. Since $T(x) = \lim_{n\to\infty} P_n(x)$, it follows that $f(x) = T(x)$ for all $|x| < 1$ iff

$$\lim_{n\to\infty} |f(x) - P_n(x)| = 0, \quad \forall\ |x| < 1. \tag{5.99}$$

(i) Use (5.58) and the integral formula (5.4) for the Taylor approximation error to show that, for any x,

$$f(x) - P_n(x) = \int_0^x \frac{(-1)^{n+2}\,(x-t)^n}{(1+t)^{n+1}}\, dt.$$

(ii) Show that, for any $0 \le x < 1$,

$$|f(x) - P_n(x)| \le \int_0^x \left(\frac{x-t}{1+t}\right)^n \frac{1}{1+t}\, dt \le x^n \ln(1+x). \tag{5.100}$$

Use (5.100) to prove that (5.99) holds for all x such that $0 \le x < 1$.

(iii) Assume that $-1 < x \leq 0$. Let $s = -x$. Show that

$$|f(x) - P_n(x)| = \int_0^s \frac{(s-z)^n}{(1-z)^{n+1}}\, dz.$$

Note that $\frac{s-z}{1-z} \leq s$, for all $0 \leq z \leq s < 1$, and obtain that

$$|f(x) - P_n(x)| \leq s^n |\ln(1-s)| = (-x)^n |\ln(1+x)|.$$

Conclude that (5.99) holds true for all x such that $-1 < x \leq 0$.

5. In the Cox-Ross-Rubinstein parametrization for a binomial tree, the up and down factors u and d, and the risk–neutral probability p of the price going up during one time step are

$$u = A + \sqrt{A^2 - 1}; \qquad (5.101)$$
$$d = A - \sqrt{A^2 - 1}; \qquad (5.102)$$
$$p = \frac{e^{r\delta t} - d}{u - d}, \qquad (5.103)$$

where

$$A = \frac{1}{2}\left(e^{-r\delta t} + e^{(r+\sigma^2)\delta t}\right).$$

Use Taylor expansions to show that, for a small time step δt, u, d and p may be approximated by

$$u = e^{\sigma\sqrt{\delta t}}; \qquad (5.104)$$
$$d = e^{-\sigma\sqrt{\delta t}}; \qquad (5.105)$$
$$p = \frac{1}{2} + \frac{1}{2}\left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\delta t}. \qquad (5.106)$$

In other words, write the Taylor expansion for (5.101–5.103) and for (5.104–5.106) and show that they are identical if all the terms of order $O(\delta t)$ and smaller are neglected.

Hint: Show that $A \approx 1 + \frac{\sigma^2 \delta t}{2}$, by using the linear Taylor expansion (5.15). Substitute this approximation in the formulas (5.101) and (5.102) and use the following linear Taylor expansions

$$\sqrt{1+x} = 1 + \frac{x}{2} + O(x^2), \quad \text{as } x \to 0;$$
$$\sqrt{1-x} = 1 - \frac{x}{2} + O(x^2), \quad \text{as } x \to 0.$$

6. (i) What is the approximate value $P_{approx,r=0,q=0}$ of an at-the-money put option on a non–dividend–paying underlying asset with spot price $S = 60$, volatility $\sigma = 0.25$, and maturity $T = 1$ year, if the constant risk–free interest rate is $r = 0$?

(ii) Compute the Black–Scholes value $P_{BS,r=0,q=0}$ of the put option, and estimate the relative approximate error

$$\frac{|P_{BS,r=0,q=0} - P_{approx,r=0,q=0}|}{P_{BS,r=0,q=0}}.$$

(iii) Assume that $r = 0.06$ and $q = 0.03$. Use formula (5.76) to compute $P_{approx,r=0.06,q=0.03}$, and estimate the relative approximate error

$$\frac{|P_{BS,r=0.06,q=0.03} - P_{approx,r=0.06,q=0.03}|}{P_{BS,r=0.06,q=0.03}}, \qquad (5.107)$$

where $P_{BS,r=0.06,q=0.03}$ is the Black–Scholes value of the put option.

7. It is interesting to note that the approximate formulas (5.75) and (5.76) for ATM call and put options do not satisfy the Put–Call parity:

$$P + Se^{-qT} - C = S\left(e^{-qT} - (r-q)T\right) \neq S\, e^{-rT} = K\, e^{-rT}.$$

Based on the linear Taylor expansion $e^{-x} \approx 1 - x$, the formulas (5.75) and (5.76) can be modified to accommodate the Put–Call parity, by replacing rT and qT by $1 - e^{-rT}$ and $1 - e^{-qT}$, respectively. The resulting formulas are

$$C \approx \sigma S \sqrt{\frac{T}{2\pi}}\, \frac{e^{-qT} + e^{-rT}}{2} + \frac{S\,(e^{-qT} - e^{-rT})}{2}; \qquad (5.108)$$
$$P \approx \sigma S \sqrt{\frac{T}{2\pi}}\, \frac{e^{-qT} + e^{-rT}}{2} - \frac{S\,(e^{-qT} - e^{-rT})}{2}. \qquad (5.109)$$

(i) Show that the Put–Call parity is satisfied by the approximations (5.108) and (5.109).

(ii) Estimate how good the new approximation (5.109) is, for an ATM put with $S = 60$, $q = 0.03$, $\sigma = 0.25$, and $T = 1$, if $r = 0.06$, by computing the corresponding relative approximate error. Compare this error with the relative approximate error (5.107) found in the previous exercise, when the approximation (5.76) was used.

8. Consider an ATM put option with strike 40 on a non–dividend paying asset with volatility 30%, and assume zero interest rates.

   Compute the relative approximation error of the approximation $P \approx \sigma S\sqrt{\frac{T}{2\pi}}$ if the put option expires in 3, 5, 10, and 20 years.

9. A five year bond worth 101 has duration 1.5 years and convexity equal to 2.5. Use both formula (2.59), which does not include any convexity adjustment, and formula (5.96) to find the price of the bond if the yield increases by ten basis points (i.e., 0.001), fifty basis points, one percent, and two percent, respectively.

Chapter 6

Finite Differences. Black–Scholes PDE.

Finite difference approximations for first order derivatives: forward, backward and central approximations. The central finite difference approximation for second order derivatives. Order of approximation.

Finite difference discretization and numerical solution of ODEs.

6.1 Forward, backward, and central finite difference approximations

Finite differences are used to approximate the values of derivatives of a given function f at a point a by values of f at a finite number of points near a.

The simplest example of a finite difference approximation is the approximation of $f'(a)$ by $f(a)$ and $f(a+h)$, where $h > 0$ is a small positive number. By definition (1.1),

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

We can therefore approximate $f'(a)$ by

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}. \qquad (6.1)$$

If $h > 0$, this is the forward finite difference approximation of $f'(a)$.

To estimate how good the forward approximation (6.1) is, assume that $f(x)$ is twice differentiable and that $f''(x)$ is a continuous function in an interval around the point a. Using the first order Taylor approximation (5.12) of $f(x)$, we find that

$$f(x) = f(a) + (x-a)f'(a) + O\left((x-a)^2\right),$$

as $x \to a$. If we let $x = a + h$, we obtain that

$$f(a+h) = f(a) + hf'(a) + O\left(h^2\right),$$

as $h \to 0$. By solving for $f'(a)$, it follows that

$$f'(a) \;=\; \frac{f(a+h) - f(a)}{h} \;+\; O(h),$$

as $h \to 0$. Thus, the forward finite difference approximation of the first derivative is a first order approximation.

The most commonly used finite difference approximations, and the corresponding orders of approximation, are presented below:

• Forward finite difference approximation for the first derivative:

Assume that $f(x)$ is twice differentiable and that $f''(x)$ is continuous in an interval around the point a. Let $h > 0$.
The forward finite difference approximation

$$f'(a) \;\approx\; \frac{f(a+h) - f(a)}{h} \tag{6.2}$$

of $f'(a)$ is a first order approximation, i.e.,

$$f'(a) \;=\; \frac{f(a+h) - f(a)}{h} \;+\; O(h), \tag{6.3}$$

as $h \to 0$.

• Backward finite difference approximation for the first derivative:

Assume that $f(x)$ is twice differentiable and that $f''(x)$ is continuous in an interval around the point a. Let $h > 0$.
The backward finite difference approximation

$$f'(a) \;\approx\; \frac{f(a) - f(a-h)}{h} \tag{6.4}$$

of $f'(a)$ is a first order approximation, i.e.,

$$f'(a) \;=\; \frac{f(a) - f(a-h)}{h} \;+\; O(h), \tag{6.5}$$

as $h \to 0$.

• Central finite difference approximation for the first derivative:

Assume that $f(x)$ is three times differentiable and that $f^{(3)}(x)$ is continuous in an interval around the point a. Let $h > 0$.
The central finite difference approximation

$$f'(a) \;\approx\; \frac{f(a+h) - f(a-h)}{2h} \tag{6.6}$$

of $f'(a)$ is a second order approximation, i.e.,

$$f'(a) \;=\; \frac{f(a+h) - f(a-h)}{2h} \;+\; O\left(h^2\right), \tag{6.7}$$

as $h \to 0$.

• Central finite difference approximation for the second derivative:

Assume that $f(x)$ is four times differentiable and that $f^{(4)}(x)$ is continuous in an interval around the point a. Let $h > 0$.
The central finite difference approximation

$$f''(a) \;\approx\; \frac{f(a+h) \;-\; 2f(a) \;+\; f(a-h)}{h^2} \tag{6.8}$$

of $f''(a)$ is a second order approximation, i.e.,

$$f''(a) \;=\; \frac{f(a+h) \;-\; 2f(a) \;+\; f(a-h)}{h^2} \;+\; O\left(h^2\right), \tag{6.9}$$

as $h \to 0$.

The finite difference approximations above can be derived using the Taylor approximation (5.7) of $f(x)$, i.e.,

$$\begin{aligned} f(x) \;=\;& f(a) \;+\; (x-a)f'(a) \;+\; \frac{(x-a)^2}{2}f''(a) \;+\; \ldots \;+\; \frac{(x-a)^n}{n!}f''(a) \\ &+ O\left((x-a)^{n+1}\right), \end{aligned} \tag{6.10}$$

as $x \to a$.

The forward finite difference approximation (6.3) was discussed in detail above. Proving that the backward finite difference approximation (6.5) is a first order approximation follows along similar lines.

To establish the central finite difference approximation (6.7) for the first derivative we use the quadratic Taylor approximation (6.10) of $f(x)$ for $n = 2$, i.e.,

$$f(x) \;=\; f(a) \;+\; (x-a)f'(a) \;+\; \frac{(x-a)^2}{2}f''(a) \;+\; O\left((x-a)^3\right), \tag{6.11}$$

as $x \to a$. By letting $x = a+h$ and $x = a-h$ in (6.11), we obtain that

$$f(a+h) \;=\; f(a) \;+\; hf'(a) \;+\; \frac{h^2}{2}f''(a) \;+\; O\left(h^3\right); \tag{6.12}$$

$$f(a-h) \;=\; f(a) \;-\; hf'(a) \;+\; \frac{h^2}{2}f''(a) \;+\; O\left(h^3\right), \tag{6.13}$$

as $h \to 0$. Subtract (6.13) from (6.12) to find that

$$f(a+h) - f(a-h) = 2hf'(a) + O\left(h^3\right), \tag{6.14}$$

as $h \to 0$, since $O\left(h^3\right) - O\left(h^3\right) = O\left(h^3\right)$; see section 0.4 for details.

We solve for $f'(a)$ in (6.14) and obtain the central finite difference approximation for the first derivative (6.7):

$$f'(a) = \frac{f(a+h) - f(a-h)}{2h} + O\left(h^2\right),$$

as $h \to 0$.

To establish the central finite difference approximation for the second derivative (6.9) we use the cubic Taylor approximation (6.10) of $f(x)$ for $n = 3$, i.e.,

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2}f''(a) + \frac{(x-a)^3}{6}f^{(3)}(a) + O\left((x-a)^4\right), \tag{6.15}$$

as $x \to a$. By letting $x = a+h$ and $x = a-h$ in (6.15), we obtain that

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \frac{h^3}{6}f^{(3)}(a) + O\left(h^4\right); \tag{6.16}$$

$$f(a-h) = f(a) - hf'(a) + \frac{h^2}{2}f''(a) - \frac{h^3}{6}f^{(3)}(a) + O\left(h^4\right), \tag{6.17}$$

as $h \to 0$. By adding (6.16) and (6.17), and using the fact that $O\left(h^4\right) + O\left(h^4\right) = O\left(h^4\right)$, we find that

$$f(a+h) + f(a-h) = 2f(a) + h^2 f''(a) + O\left(h^4\right), \tag{6.18}$$

as $h \to 0$. The central finite difference approximation for the second derivative (6.9) is obtained by solving for $f''(a)$ in (6.18), i.e.,

$$f''(a) = \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} + O\left(h^2\right).$$

6.2 Finite difference discretization and solution for first and second order ODEs

Consider the following ordinary differential equation (ODE) on a bounded interval: Find $y : [a,b] \to \mathbb{R}$ such that

$$y'(x) = f(x, y(x)), \quad \forall\, x \in [a,b]; \tag{6.19}$$

$$y(a) = c, \tag{6.20}$$

where f is a continuous function. The ODE is made of the differential equation (6.19) and the boundary condition (6.20).

We are not looking for a closed form solution of the ODE (6.19–6.20), which may, in fact, not exist, nor to establish the existence and uniqueness of its solution. Our goal is to find a numerical solution of the ODE (6.19–6.20) using finite differences.

The general idea of numerical solutions for ODEs is given in a nutshell below:

Solving a differential equation approximately means discretizing the computational domain where the equation must be solved by choosing a finite number of points in the domain (usually equidistant), and finding numerical values at these discrete points (nodes) that approximate the values of the exact solution of the differential equation at those points.

We partition the interval $[a,b]$ into n intervals of equal size $h = \frac{b-a}{n}$, using the nodes $x_i = a + ih$, $i = 0 : n$, i.e.,

$$x_0 = a; \; x_1 = a + h; \; \ldots \; x_{n-1} = a + (n-1)h; \; x_n = a + nh = b. \tag{6.21}$$

Finding an approximate solution of the exact solution $y(x)$ of an ODE on the computational interval $[a,b]$ discretized by using the nodes $x_0, x_1, \ldots, x_n$ given by (6.21) requires finding values $y_0, y_1, \ldots, y_n$ such that

$$\begin{array}{lll} y_0 & \text{is the approximate value of} & y(x_0); \\ y_1 & \text{is the approximate value of} & y(x_1); \\ \vdots & \vdots & \vdots \\ y_n & \text{is the approximate value of} & y(x_n). \end{array}$$

Until now, we only defined what finding a numerical solution of the ODE means, without actually specifying how the values $y_0, y_1, \ldots, y_n$ should be found in order to approximate $y(x)$ correctly. It stands to reason that this must be related to the differential equation (6.19), and this is indeed the case: we will discretize (6.19) using forward finite difference approximations for $y'(x)$; cf. (6.3). The resulting method is called Euler's method.

From (6.19), we know that $y'(x) = f(x, y(x))$ for any point $x \in [a,b]$. We only use this relationship for the discrete nodes $x_0, x_1, \ldots, x_{n-1}$, i.e.,

$$y'(x_i) = f(x_i, y(x_i)), \quad \forall\, i = 0 : (n-1). \tag{6.22}$$

Note that the function $y(x)$ from (6.22) is the *exact solution* of the differential equation (6.19).

The forward finite difference approximation of $y'(x)$ at a node x_i can be written in terms of $y(x_i)$ and $y(x_{i+1})$ as

$$y'(x_i) = \frac{y(x_{i+1}) - y(x_i)}{h} + O(h), \tag{6.23}$$

as $h \to 0$; cf. (6.3) for $f(x) = y(x)$, $a = x_i$, and $h = x_{i+1} - x_i$.

By substituting the approximation (6.23) into (6.19), we obtain that

$$\frac{y(x_{i+1}) - y(x_i)}{h} + O(h) = f(x_i, y(x_i)), \quad \forall\, i = 0 : (n-1). \tag{6.24}$$

The discretization of the ODE (6.19) is obtained from (6.24) by ignoring the $O(h)$ term and substituting the approximate values y_i for the exact values $y(x_i)$. We find that

$$\frac{y_{i+1} - y_i}{h} = f(x_i, y(x_i)), \quad \forall\, i = 0 : (n-1),$$

which can be written as

$$y_{i+1} = y_i + hf(x_i, y(x_i)), \quad \forall\, i = 0 : (n-1). \tag{6.25}$$

From the boundary condition (6.20), it follows that $y(x_0) = y(a) = c$. Therefore, we choose $y_0 = c$. Every approximate value y_i, $i = 1 : n$, can then be found recursively from (6.25). The values $(y_i)_{i=0:n}$ represent the finite difference solution of the ODE (6.19–6.20), corresponding to a uniform discretization with $n+1$ nodes.

For a finite difference scheme to be useful, it must converge and do so fast. For $i = 0 : n$, let $e_i = y_i - y(x_i)$ be the approximation error of the finite difference solution at the node x_i. We say that the finite difference scheme is convergent if and only if the maximum approximation error goes to 0 as the number of nodes n in the discretization scheme goes to infinity, i.e.,

$$\lim_{n\to\infty} \max_{i=0:n} |e_i| = \lim_{n\to\infty} \max_{i=0:n} |y_i - y(x_i)| = 0;$$

for more details on finite difference schemes and their properties, we refer the reader to the references at the end of the chapter.

For clarification purposes, we present the finite difference solution for several ODEs. Rather than using the general recursion formula (6.25), we go through the entire finite difference discretization process for each ODE.

Example: Solve the following second order linear ODE with constant coefficients using the finite difference method:
Find $y(x)$ on the interval $[0, 1]$ such that

$$y''(x) = y(x), \quad \forall\, x \in [0, 1]; \tag{6.26}$$
$$y(0) = 1; \quad y(1) = e. \tag{6.27}$$

Solution: We first note[1] that this ODE has the exact solution $y(x) = e^x$.

For the finite difference solution, we discretize the interval $[0, 1]$ by choosing the nodes $x_i = ih$, $i = 0 : n$, with $h = \frac{1}{n}$. We look for $y_0, y_1, \ldots, y_n$ such that y_i is an approximate value of $y(x_i)$, for all $i = 0 : n$, where $y(x)$ is the exact solution of the ODE (6.26–6.27).

By writing (6.26) at each interior node x_i, $i = 1 : n-1$, we find that

$$y''(x_i) = y(x_i), \quad \forall\, i = 1 : (n-1). \tag{6.28}$$

We approximate $y''(x_i)$ by using central finite differences (6.9), i.e.,

$$\begin{aligned} y''(x_i) &= \frac{y(x_i + h) - 2y(x_i) + y(x_i - h)}{h^2} + O(h^2) \\ &= \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1})}{h^2} + O(h^2), \end{aligned} \tag{6.29}$$

since $x_{i+1} = x_i + h$ and $x_{i-1} = x_i - h$.

We substitute the approximation (6.29) into (6.28) and obtain

$$\frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1})}{h^2} + O(h^2) = y(x_i), \quad \forall\, i = 1 : (n-1). \tag{6.30}$$

The discretization of the ODE (6.26) is obtained by substituting in (6.30) the approximate values y_i for the exact values $y(x_i)$, for all $i = 0 : n$, and ignoring the $O(h^2)$ term. We find that

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = y_i, \quad \forall\, i = 1 : (n-1),$$

which can be written as

$$-\,y_{i-1} + (2 + h^2)y_i - y_{i+1} = 0, \quad \forall\, i = 1 : (n-1). \tag{6.31}$$

Given the boundary conditions (6.27), i.e., $y(0) = 1$ and $y(1) = e$, we choose $y_0 = 1$ and $y_n = e$. Then, formulas (6.31) can be written in matrix form as

$$A\,Y = b, \tag{6.32}$$

[1] We also note that the ODE (6.26) can be written in a form similar to (6.19) as follows: Define $Y : [0, 1] \to \mathbb{R}$ as

$$Y(x) = \begin{pmatrix} y(x) \\ y'(x) \end{pmatrix}.$$

Then,

$$Y'(x) = \begin{pmatrix} y'(x) \\ y''(x) \end{pmatrix} = \begin{pmatrix} y'(x) \\ y(x) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} Y(x) = f(x, Y(x)),$$

where $f : [0, 1] \times \mathbb{R}^2 \to \mathbb{R}^2$ is given by $f(x, y_1, y_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$.

where A is the following $(n-1)\times(n-1)$ matrix

$$A = \begin{pmatrix} 2+h^2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2+h^2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2+h^2 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2+h^2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2+h^2 \end{pmatrix},$$

and Y and b are the following column vectors of size $n-1$:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-2} \\ y_{n-1} \end{pmatrix}; \quad b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ e \end{pmatrix}.$$

The matrix A is strictly diagonally dominated[2] and therefore all the eigenvalues of A are nonzero. Thus, A is nonsingular and the linear system (6.32) has a unique solution for any h, i.e., for any n.

Also, note that the matrix A is symmetric, banded (tridiagonal), and positive definite (all the eigenvalues of A are positive, since A is strictly diagonally dominated and its entries on the main diagonal are positive). Then, A is a symmetric positive definite matrix and the system (6.32) can be solved very efficiently using the Cholesky decomposition for the matrix A. This method requires only $O(n)$ operations to solve the linear system.

For more details on the numerical linear algebra notions discussed above, as well as on solving linear systems, we refer the reader to [27].

Numerical Solution: For $n = 8$ intervals, the numerical solution to the finite difference discretization (6.31) of the ODE (6.26) is given below:

We double the number of discretization intervals from $n = 8$ intervals up to $n = 1024$ intervals, and solve the corresponding finite difference discretization. We report the maximum approximation error

$$E(n) = \max_{i=0:n} |e_i| = \max_{i=0:n} |y_i - y(x_i)|.$$

[2]The matrix $M = (M(i,j))_{1\leq i,j\leq n}$ is strictly diagonally dominated if and only if

$$|M(k,k)| > \sum_{j=1}^{k-1} |M(k,j)| + \sum_{j=k+1}^{n} |M(k,j)|, \quad \forall\, k = 1:n.$$

i	x_i	y_i	$y(x_i)$	$\lvert y_i - y(x_i)\rvert$
1	0.125	1.133244774	1.133148453	0.000096321
2	0.250	1.284196498	1.284025417	0.000171081
3	0.375	1.455213792	1.454991415	0.000222377
4	0.500	1.648968801	1.648721271	0.000247530
5	0.625	1.868488948	1.868245957	0.000242990
6	0.750	2.117204235	2.117000017	0.000204218
7	0.875	2.399000837	2.398875294	0.000125544

n	$E(n)$	$E(n)/E(2n)$
8	0.000247530	3.9669
16	0.000062399	3.9969
32	0.000015612	3.9966
64	0.000003906	3.9999
128	0.000000977	4.0000
256	0.000000244	3.9999
512	0.000000061	4.0000
1024	0.000000015	

Note that the maximum approximation error decreases by a factor of approximately 4 every time the number of intervals doubles. The finite difference discretization scheme is (quadratically) convergent.

Example: Solve the following ODE using the finite difference method: Find $y(x)$ on the interval $[-1,1]$ such that

$$y''(x) + 3y'(x) + 2y(x) = 0, \quad \forall\, x \in [-1,1]; \tag{6.33}$$

$$y(-1) = 1 + e^4; \quad y(1) = 1 + \frac{1}{e^2}. \tag{6.34}$$

Solution: The roots of the characteristic polynomial $z^2 + 3z + 2 = 0$ corresponding to the ODE (6.33) are -1 and -2. Therefore, the general formula for the solution of the ODE is

$$y(x) = C_1 e^{-x} + C_2 e^{-2x},$$

For the boundary conditions (6.34), the solution of the ODE (6.33) is

$$y(x) = e^{-x-1} + e^{-2x+2}.$$

To find a numerical solution of the ODE (6.33), we discretize the interval $[-1, 1]$ using the nodes

$$x_0 = -1;\ x_1 = -1 + h;\ \ldots\ x_{n-1} = -1 + (n-1)h;\ x_n = -1 + nh = 1,$$

which partition the interval $[-1, 1]$ into n intervals of equal size $h = \frac{2}{n}$.

We write the ODE (6.33) at each interior node x_i, $i = 1 : n-1$, to obtain

$$y''(x_i) \ + \ 3y'(x_i) \ + \ 2y(x_i) \ = \ 0, \quad \forall\, i = 1 : (n-1). \tag{6.35}$$

We need to choose finite difference approximations for both $y''(x_i)$ and $y'(x_i)$.

The finite difference approximations of every derivative must be of the same order. The resulting finite difference scheme for the ODE is only as good as the lowest order approximation of any derivative.

We use the second order central difference approximations (6.9) and (6.7) for $y''(x_i)$ and $y'(x_i)$, respectively, i.e.,

$$y''(x_i) \ = \ \frac{y(x_{i+1}) \ - \ 2y(x_i) \ + \ y(x_{i-1})}{h^2} \ + \ O(h^2); \tag{6.36}$$

$$y'(x_i) \ = \ \frac{y(x_{i+1}) \ - \ y(x_{i-1})}{2h} \ + \ O(h^2). \tag{6.37}$$

We substitute (6.36) and (6.37) into (6.35), use the approximate values y_i instead of the exact values $y(x_i)$, and ignore the $O(h^2)$ term. We obtain the following second order finite difference discretization of (6.35):

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \ + \ 3\,\frac{y_{i+1} - y_{i-1}}{2h} \ + \ 2y_i \ = \ 0.$$

After multiplication by h^2, we find that

$$y_{i+1} - 2y_i + y_{i-1} \ + \ \frac{3h}{2}\,(y_{i+1} - y_{i-1}) \ + \ 2h^2 y_i \ = \ 0,$$

which can be written as

$$-\left(1 - \frac{3h}{2}\right) y_{i-1} + 2(1-h^2)y_i - \left(1 + \frac{3h}{2}\right) y_{i+1} = 0, \quad \forall\, i = 1 : (n-1). \tag{6.38}$$

Given the boundary conditions (6.34), we choose $y_0 = 1 + e^4$ and $y_n = 1 + \frac{1}{e^2}$. The equations (6.38) can be written in matrix form as

$$A\,Y \ = \ b, \tag{6.39}$$

where A is a tridiagonal $(n-1) \times (n-1)$ matrix given by

$$\begin{aligned} A(i,i) \ &= \ 2(1-h^2), \quad \forall\ i = 1 : (n-1); \\ A(i,i+1) \ &= \ -\left(1 + \tfrac{3h}{2}\right), \quad \forall\ i = 1 : (n-2); \\ A(i,i-1) \ &= \ -\left(1 - \tfrac{3h}{2}\right), \quad \forall\ i = 2 : (n-1), \end{aligned}$$

and Y and b are the following column vectors of size $n-1$:

$$Y \ = \ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-2} \\ y_{n-1} \end{pmatrix}; \quad b \ = \ \begin{pmatrix} \left(1 - \frac{3h}{2}\right)\left(1 + e^4\right) \\ 0 \\ \vdots \\ 0 \\ \left(1 + \frac{3h}{2}\right)\left(1 + \frac{1}{e^2}\right) \end{pmatrix}.$$

It is not a priori clear whether the matrix A is non-singular, and therefore whether the system (6.39) has a unique solution. However, this is a non-relevant question to ask when solving linear systems numerically, e.g., by using the LU decomposition (with pivoting) of the matrix A. If the matrix A is singular, then the LU decomposition will fail and no solution will be computed. Otherwise, if A is nonsingular, the unique solution of the linear system (6.39) will be computed in $O(n)$ operations. For more details on numerical linear algebra topics, including LU and Cholesky decompositions and the numerical solution of linear systems, we again refer the reader to [27].

Numerical Solution: For $n = 8$ intervals, the numerical solution to the finite difference discretization (6.38) of the ODE (6.33) is given below:

i	x_i	y_i	$y(x_i)$	$\lvert y_i - y(x_i)\rvert$
1	0.125	32.920620220	33.894252742	0.973632522
2	0.250	19.619868466	20.692067583	1.072199117
3	0.375	11.790447809	12.654860513	0.864412704
4	0.500	7.159761345	7.756935540	0.597174195
5	0.625	4.404016467	4.768193867	0.364177400
6	0.750	2.751040025	2.941411989	0.190371963
7	0.875	1.749592549	1.822495214	0.072902667

We double the number of discretization intervals from $n = 8$ intervals up to $n = 1024$ intervals, and report the maximum approximation error $E(n) = \max_{i=0:n} |y_i - y(x_i)|$.

The maximum approximation error decreases by a factor of approximately 4, when the number of intervals doubles. The finite difference discretization scheme is (quadratically) convergent.

Example: Solve the following ODE using the finite difference method: Find $y(x)$ on the interval $[0, 1]$ such that

$$(1+x)y''(x) \ + \ xy'(x) \ - \ 2y(x) \ = \ 0, \quad \forall\, x \in [0,1]; \tag{6.40}$$

$$y(0) \ = \ 1; \quad y(1) \ = \ 1. \tag{6.41}$$

n	$E(n)$	$E(n)/E(2n)$
8	1.072199117	4.1851
16	0.256195595	4.0237
32	0.063672030	4.0132
64	0.015865729	4.0024
128	0.003964076	4.0006
256	0.000990877	4.0001
512	0.000247710	4.0000
1024	0.000061927	

Solution: We discretize the interval $[0,1]$ by choosing the nodes $x_i = ih$, $i = 0 : n$, where $h = \frac{1}{n}$. We look for $y_0, y_1, \ldots, y_n$ such that y_i is an approximate value of $y(x_i)$, for all $i = 0 : n$.

By writing the ODE (6.40) at each interior node $x_i = ih$, we obtain

$$(1 + x_i)y''(x_i) \ + \ x_i y'(x_i) \ - \ 2y(x_i) \ = \ 0, \quad \forall\, i = 1 : (n-1). \tag{6.42}$$

As in the previous example, we use the central difference approximations (6.9) and (6.7) for $y''(x_i)$ and $y'(x_i)$, respectively, i.e.,

$$y''(x_i) \ = \ \frac{y(x_{i+1}) \ - \ 2y(x_i) \ + \ y(x_{i-1})}{h^2} \ + \ O(h^2); \tag{6.43}$$

$$y'(x_i) \ = \ \frac{y(x_{i+1}) \ - \ y(x_{i-1})}{2h} \ + \ O(h^2). \tag{6.44}$$

We substitute (6.43) and (6.44) into (6.42). Using the approximate values y_i for the exact values $y(x_i)$, for $i = 0 : n$, and ignoring the $O(h^2)$ term, we obtain the following second order finite difference discretization of (6.40):

$$(1 + ih)\,\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \ + \ ih\,\frac{y_{i+1} - y_{i-1}}{2h} \ - \ 2y_i \ = \ 0,$$

since $x_i = ih$. After multiplication by $-h^2$, we find that, for any $i = 1 : (n-1)$,

$$-\left(1 + ih - \frac{ih^2}{2}\right) y_{i-1} + 2(1+ih+h^2)y_i - \left(1 + ih + \frac{ih^2}{2}\right) y_{i+1} \ = \ 0. \tag{6.45}$$

Given the boundary conditions (6.41), we choose $y_0 = 1$ and $y_n = 1$. The equations (6.45) can be written in matrix form as

$$A\,Y \ = \ b,$$

where A is a tridiagonal $(n-1) \times (n-1)$ matrix given by

$$\begin{aligned}
A(i,i) \ &= \ 2(1 + ih + h^2), && \forall\ i = 1 : (n-1);\\
A(i,i+1) \ &= \ -\left(1 + ih + \tfrac{ih^2}{2}\right), && \forall\ i = 1 : (n-2);\\
A(i,i-1) \ &= \ -\left(1 + ih - \tfrac{ih^2}{2}\right), && \forall\ i = 2 : (n-1),
\end{aligned}$$

and Y and b are the following column vectors of size $n-1$

$$Y \ = \ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-2} \\ y_{n-1} \end{pmatrix}; \quad b \ = \ \begin{pmatrix} 1 + h - \frac{h^2}{2} \\ 0 \\ \vdots \\ 0 \\ 1 + (n-1)h + \frac{(n-1)h^2}{2} \end{pmatrix}.$$

Numerical Solution: For $n = 8$ intervals, the numerical solution to the finite difference discretization (6.45) of the ODE (6.40) is given below.

i	x_i	y_i
1	0.125	0.927324726
2	0.250	0.881233241
3	0.375	0.858038663
4	0.500	0.854795605
5	0.625	0.869129726
6	0.750	0.899112567
7	0.875	0.943167827

Since the exact solution $y(x)$ of the ODE (6.40–6.41) is not provided, the approximation errors of the finite difference solution can no longer be computed as $e_i = |y_i - y(x_i)|$, $i = 0 : n$.

We double the number of intervals from $n = 8$ intervals up to $n = 1024$ intervals. We investigate the convergence patterns of the numerical solution at the points $\{\frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}\}$, which are nodes in the discretization of the interval $[0,1]$ for every value of n considered here.

Let $n = 2^j$, with $3 \leq j \leq 10$, and denote by y_i the approximation to the exact solution of the ODE evaluated at the node $x_i = \frac{i}{n}$. Note that x_i is one of the nodes $\frac{k}{8}$, with $k = 1 : 7$, if and only if $i = \frac{kn}{8}$. Thus, for each n, we report the approximate values $y_{kn/8}$, for $k = 1 : 7$.

As the number of intervals increases, the pattern exhibited by the numerical solution suggests that the finite difference discretization scheme is convergent.

n	$y_{n/8}$	$y_{2n/8}$	$y_{3n/8}$	$y_{4n/8}$
8	0.927324726	0.881233241	0.858038663	0.854795605
16	0.927230008	0.881096027	0.857892126	0.854660722
32	0.927206081	0.881061382	0.857855140	0.854626687
64	0.927200084	0.881052699	0.857845871	0.854618159
128	0.927198583	0.881050527	0.857843553	0.854616026
256	0.927198208	0.881049984	0.857842973	0.854615492
512	0.927198114	0.881049848	0.857842828	0.854615359
1024	0.927198091	0.881049814	0.857842792	0.854615325

n	$y_{5n/8}$	$y_{6n/8}$	$y_{7n/8}$
8	0.869129726	0.899112567	0.943167827
16	0.869019591	0.899035166	0.943127841
32	0.868991807	0.899015643	0.943117757
64	0.868984845	0.899010752	0.943115230
128	0.868983104	0.899009528	0.943114598
256	0.868982669	0.899009222	0.943114440
512	0.868982560	0.899009146	0.943114401
1024	0.868982533	0.899009127	0.943114391

FINANCIAL APPLICATIONS

Finite difference approximations for the Greeks.

The Black–Scholes PDE.

Connections between the Greeks derived from the Black–Scholes PDE.

6.3 Finite difference approximations for the Greeks

The reason we were able to compute exact formulas for the Greeks of European plain vanilla options is that closed formulas (the Black–Scholes formulas) exist for pricing those options; see section 3.6. If a closed formula for the price V of a derivative security does not exist (which is almost always the case), we cannot compute closed formulas for the partial derivatives of V, and therefore for the Greeks of the derivative security. In this case, finite differences are used to obtain approximate values for the Greeks.

Using the forward and central finite difference approximations (6.3) and

(6.7), the Delta of a derivative security can be approximated as follows:

$$\Delta(V) \approx \frac{V(S+dS)-V(S)}{dS};$$
$$\Delta(V) \approx \frac{V(S+dS)-V(S-dS)}{2dS},$$

where, e.g., $V(S+dS)$ is the value of the derivative security when the spot price of the underlying asset is $S+dS$.

The Gamma of a derivative security can be approximated using the central difference approximation (6.9), i.e.,

$$\Gamma(V) \approx \frac{V(S+dS)-2V(S)+V(S-dS)}{(dS)^2}.$$

For the other Greeks, i.e., ρ, Θ, and vega, the forward difference (6.3) is the most commonly used approximation formula. Thus,

$$\text{vega}(V) \approx \frac{V(\sigma+d\sigma)-V(\sigma)}{d\sigma};$$
$$\rho(V) \approx \frac{V(r+dr)-V(r)}{dr};$$
$$\Theta(V) \approx \frac{V(t+dt)-V(t)}{dt}.$$

We note that $\Theta(V)$ can also be computed as

$$\Theta(V) = -\frac{\partial V}{\partial T};$$

see (7.24) and section 7.4 for more details. Then, a finite difference approximation for Θ can be obtained as follows:

$$\Theta(V) \approx -\frac{V(T+dT)-V(T)}{dT}. \tag{6.46}$$

6.4 The Black–Scholes PDE

If the underlying asset follows a lognormal process, the value $V(S,t)$ of a plain vanilla European option satisfies a partial differential equation, i.e., an equation involving partial derivatives of V with respect to the spot price S of the underlying asset and the time t. This partial differential equation is called the Black–Scholes PDE.

More formally, let T be the maturity of the plain vanilla European option, and let σ and q be the volatility and continuous dividend rate of the underlying asset. Using, e.g., the hedging portfolio[3] considered in section 5.4, it can be shown that $V(S,t)$ satisfies the Black–Scholes PDE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV = 0, \tag{6.47}$$

for all $S > 0$ and $0 < t < T$. The value $V(S,T)$ of the option at maturity is equal to the payoff of the option at maturity, i.e.,

$$V(S,T) = \max(S-K,0), \quad \text{for call options;} \tag{6.48}$$
$$V(S,T) = \max(K-S,0), \quad \text{for put options.} \tag{6.49}$$

The solution to the Black–Scholes PDE (6.47) with boundary conditions (6.48) is the Black–Scholes value (3.57) of a call option. Similarly, the solution to the Black–Scholes PDE (6.47) with boundary conditions (6.49) is the Black–Scholes value (3.58) of a put option.

The Black–Scholes PDE can also be written in term of the differential operator L_{BS}, i.e., the Black–Scholes operator,

$$L_{BS} = \frac{\partial}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2} + (r-q)S\frac{\partial}{\partial S} - r \tag{6.50}$$

as

$$L_{BS}V = 0, \ \forall\, S > 0, \ \forall\, 0 < t < T.$$

This Black–Scholes operator L_{BS} has the following properties:

- L_{BS} is a *linear* operator, i.e.,

$$L_{BS}(V_1) + L_{BS}(V_2) = L_{BS}(V_1 + V_2);$$
$$L_{BS}(cV) = cL_{BS}(V), \ \forall\, c \in \mathbb{R};$$

- L_{BS} is a second order parabolic operator, i.e., the highest order of the partial derivatives in t is $\frac{\partial V}{\partial t}$, and the highest order of the partial derivatives in S is $\frac{\partial^2 V}{\partial S^2}$;
- L_{BS} has nonconstant coefficients;
- L_{BS} is a backward operator in time, i.e., the coefficients of $\frac{\partial V}{\partial t}$ and $\frac{\partial^2 V}{\partial S^2}$ have the same sign.

We note the values of many financial instruments other than the plain vanilla European options also satisfy the Black–Scholes PDE. For example, if

[3]The hedging portfolio is made of a long position in the option and a short position in Δ units of the underlying asset.

the underlying asset does not pay dividends, the Black–Scholes operator can be written as

$$L_{BS}V = \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial}{\partial S} - rV;$$

cf. (6.50) for $q = 0$. Then, the value $V(S,t) = S$ of the non–dividend–paying underlying asset satisfies the Black–Scholes PDE, since

$$\frac{\partial V}{\partial t} = 0; \quad \frac{\partial V}{\partial S} = 1; \quad \text{and} \quad \frac{\partial^2 V}{\partial S^2} = 0,$$

and therefore

$$L_{BS}V = 0 + \frac{1}{2}\sigma^2 S^2 \cdot 0 + rS \cdot 1 - rS = 0.$$

However, this does not mean that the Black–Scholes values of plain vanilla options are not unique, since the function $V(S,t) = S$ does not satisfy either one of the boundary conditions (6.48) or (6.49). The only solutions to the Black–Scholes PDE that satisfy the boundary conditions (6.48) or (6.49), respectively, correspond to the Black–Scholes formulas for a call or a put option, respectively.

6.4.1 Financial interpretation of the Black–Scholes PDE

To find the financial meaning of the terms from the Black–Scholes PDE (6.47), we rewrite it using the Delta and Gamma of the option as

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \Gamma + (r-q)\Delta S - rV = 0,$$

where $\Delta = \frac{\partial V}{\partial S}$ and $\Gamma = \frac{\partial^2 V}{\partial S^2}$. Then,

$$\frac{\partial V}{\partial t} = r(V - \Delta S) - \frac{\sigma^2 S^2}{2}\Gamma + q\Delta S. \tag{6.51}$$

The time discretization of (6.51) between time t and time $t + dt$ is obtained by substituting the forward finite difference approximation $\frac{V(S,t+dt)-V(S,t)}{dt}$ for $\frac{\partial V}{\partial t}$ in (6.51). Then,

$$\frac{V(S,t+dt) - V(S,t)}{dt} = r(V - \Delta S) - \frac{\sigma^2 S^2}{2}\Gamma + q\Delta S,$$

which can be written as

$$dV = r(V - \Delta S)\,dt - \frac{\sigma^2 S^2}{2}\Gamma dt + q\Delta S dt, \tag{6.52}$$

where $dV = V(S, t + dt) - V(S, t)$ and $V = V(S, t)$ in the right hand side of (6.52).

We note that a plain vanilla European option with price V can be replicated dynamically using a portfolio made of a long position in Δ units of the underlying asset and a cash position of size $V - \Delta S$.

Formula (6.52) can then be interpreted as follows:

The change dV in the price of the option between time t and time $t + dt$ consists of

- the interest $r(V - \Delta S)dt = r\left(V - S\frac{\partial V}{\partial S}\right) dt$ realized on the cash position from the replicating portfolio;
- the hedging costs $-\frac{\sigma^2 S^2}{2}\Gamma dt = -\frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2}dt$ (also called slippage);
- the dividend gain $q\Delta S dt = qS\frac{\partial V}{\partial S}dt$ of the Δ position in the underlying asset.

6.4.2 Connections between the Greeks derived from the Black–Scholes PDE

Recall from section 3.6 that the Greeks represent the rates of change of the price of a derivative security with respect to various parameters. For example, Δ and Θ are the rates of change of the price of the derivative security with respect to the spot price S of the underlying asset, and with respect to time t, respectively, while Γ is the rate of change of Δ with respect to S, i.e.,

$$\Delta = \frac{\partial V}{\partial S}; \quad \Gamma = \frac{\partial^2 V}{\partial S^2}; \quad \Theta = \frac{\partial V}{\partial t}.$$

The Black–Scholes PDE (6.47), i.e.,

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - q)S\frac{\partial V}{\partial S} - rV = 0,$$

can be written in terms of the Greeks as

$$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma + (r - q)S\Delta - rV = 0,$$

which is equivalent to

$$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma = r(V - \Delta S) + q\Delta S. \tag{6.53}$$

In general, the right hand side of (6.53) is small compared to the terms Θ and $\sigma^2 S^2 \Gamma$. This means that, for most instances

$$1 + \frac{\sigma^2 S^2}{2} \cdot \frac{\Gamma}{\Theta} \approx 0.$$

Therefore Θ and Γ have, for most cases, different signs. Moreover, if Θ is large in absolute value, then Γ is also large in absolute value, and vice versa.

Plain vanilla European call and put options have positive Γ, since Δ is an increasing function of the price of the underlying asset. This means that, in general, Θ is negative: as the option moves closer to maturity its value decreases (all other things such as the spot price being assumed not to change), since the potential for a larger payoff at maturity, or for the option to expire in the money, decreases.

We note that one instance when this connection between Θ and Γ does not hold is for European put options that are deep in the money, when both Θ and Γ are positive.

Example: If interest rates are constant at 5%, the Gamma and Theta of a six months at–money–put option on a non–dividend–paying asset with spot price 40 and volatility 25% are

$$\Gamma = 0.054949 \quad \text{and} \quad \Theta = -1.813373,$$

and Γ and Θ have different signs.

The Gamma and Theta of a six month put with strike 50 on the same underlying asset, i.e., of a put which is twenty percent in the money, are

$$\Gamma = 0.033109 \quad \text{and} \quad \Theta = 0.054948.$$

In other words, Θ is positive and Γ and Θ have the same sign.

6.5 References

Finite difference approximations and solutions of ODEs are fundamental topics in numerical analysis. A good treatment of finite difference methods, including theoretical convergence proofs, can be found in Iserles [15], while more advanced topics are covered in Strikwerda [30].

Applications of finite differences for financial applications are included in Achdou and Pirroneau [2] and Tavella [32]. A monograph on finite difference methods in financial engineering was written by Duffy [8].

6.6 Exercises

1. A butterfly spread is made of a long position in a call option with strike $K - x$, a long position in a call option with strike $K + x$, and a short position in two calls with strike K. The options are on the same underlying asset and have the same maturities.

   (i) Show that the value of the butterfly spread is

   $$C(K+x) - 2C(K) + C(K-x),$$

   where, e.g., $C(K+x)$ denotes the price of the call with strike $K + x$.

   (ii) Show that, in the limiting case when x goes to 0, the **value** of a position in $\frac{1}{x^2}$ butterfly spreads as above converges to the second order partial derivative of the value of the option, C, with respect to strike K, i.e., show that

   $$\lim_{x \searrow 0} \frac{C(K+x) - 2C(K) + C(K-x)}{x^2} = \frac{\partial^2 C}{\partial K^2}(K).$$

   (iii) Show that, in the limiting case when $x \to 0$, the **payoff at maturity** of a position in $\frac{1}{x}$ butterfly spreads as above is going to approximate the payoff of a derivative security that pays 1 if the underlying asset expires at K, and 0 otherwise.

   Note: A security that pays 1 in a certain state and 0 is any other state is called an Arrow-Debreu security, and its price is called the Arrow-Debreu price of that state. A position in $\frac{1}{x}$ butterfly spreads as above, with x small, is a synthetic way to construct an Arrow-Debreu security for the state $S(T) = K$.

2. A bullspread is made of a long position in a call option with strike K and a short position in a call option with strike $K + x$, both options being on the same underlying asset and having the same maturities. Let $C(K)$ and $C(K+x)$ be the values (at time t) of the call options with strikes K and $K + x$, respectively.

   (i) The **value** of a position in $\frac{1}{x}$ bull spreads is $\frac{C(K)-C(K+x)}{x}$. In the limiting case when x goes to 0, show that

   $$\lim_{x \searrow 0} \frac{C(K) - C(K+x)}{x} = -\frac{\partial C}{\partial K}(K).$$

   (ii) Show that, in the limiting case when $x \to 0$, the **payoff at maturity** of a position in $\frac{1}{x}$ bullspreads as above is going to approximate the payoff of a derivative security that pays 1 if the price of the underlying asset at expiry is above K, and 0 otherwise.

   Note: A position in $\frac{1}{x}$ bullspreads as above, with x small, is a synthetic way to construct a cash–or–nothing call maturing at time T.

3. Find a second order finite difference approximation for $f'(a)$ using $f(a)$, $f(a+h)$, and $f(a+2h)$.

   Note: This type of approximation is needed, e.g., when discretizing a PDE with boundary conditions involving derivatives of the solution (also called Robin boundary conditions). For example, for Asian Options (continuous computed average rate call, to be more precise), this type of finite difference approximation is used to discretize the boundary condition $\frac{\partial H}{\partial t} + \frac{\partial H}{\partial R} = 0$, for $R = 0$.

4. Find a central finite difference approximation for the fourth derivative of f at a, i.e., for $f^{(4)}(a)$, using $f(a-2h)$, $f(a-h)$, $f(a)$, $f(a+h)$, and $f(a+2h)$. What is the order of this finite difference approximation?

5. The goal of this exercise is to emphasize the importance of symmetry in finite difference approximations. Recall from (6.7) and (6.9) that the central difference approximations for the first and second order derivatives are

   $$\begin{aligned} f'(a) &= \frac{f(a+h) - f(a-h)}{2h} + O\left(h^2\right); \\ f''(a) &= \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} + O\left(h^2\right), \end{aligned}$$

   as $h \to 0$. In other words, $f'(a)$ and $f''(a)$ are approximated to second order accuracy by using the value of f at the point a and at the points $a - h$ and $a + h$ that are symmetric with respect to a.

   We investigate what happens if symmetry is not required.

   (i) Find a second order finite difference approximation of $f'(a)$ using $f(a)$, $f(a-h)$ and $f(a+2h)$.

   (ii) Find a first order finite difference approximation of $f''(a)$ using $f(a)$, $f(a-h)$ and $f(a+2h)$. Note that, in general, a second order finite difference approximation of $f''(a)$ using $f(a)$, $f(a-h)$ and $f(a+2h)$ does not exist.

   Let $\beta < a < \gamma$ such that $a - \beta = C(\gamma - a)$, where C is a constant.

(iii) Find a finite difference approximation of $f'(a)$ using $f(a)$, $f(\beta)$, and $f(\gamma)$, which is second order in terms of $|\gamma - a|$, i.e., where the residual term is $O\left(|\gamma - a|^2\right)$.

(iv) Find a finite difference approximation of $f''(a)$ using $f(a)$, $f(\beta)$, and $f(\gamma)$, which is first order in terms of $|\gamma - a|$. Show that, in general, a second order finite difference approximation of $f''(a)$ using $f(a)$, $f(\beta)$ and $f(\gamma)$ is not possible, unless $a = \frac{\beta+\gamma}{2}$, i.e., unless β and γ are symmetric with respect to a.

6. Consider the following first order ODE:

$$\begin{aligned} y'(x) &= y(x), \quad \forall\, x \in [0,1]; \\ y(0) &= 1. \end{aligned}$$

(i) Discretize the interval $[0,1]$ using the nodes $x_i = ih$, $i = 0 : n$, where $h = \frac{1}{n}$. Use forward finite differences to obtain the following finite difference discretization of the ODE:

$$y_{i+1} = (1+h)y_i, \quad \forall\, i = 0 : (n-1),$$

with $y_0 = 1$. Show that

$$y_i = (1+h)^i, \quad \forall\, i = 0 : n.$$

(ii) Note that $y(x) = e^x$ it the exact solution of the ODE. Let

$$e_i = y_i - y(x_i) = (1+h)^i - e^{ih}$$

be the approximation error of the finite difference solution at the node x_i, $i = 0 : n$. Our goal is to show that this finite difference discretization is convergent, i.e., that

$$\lim_{n\to\infty} \max_{i=0:n} |e_i| = 0.$$

It is easy to see that e_i can also be written as

$$e_i = e^{i\ln(1+h)} - e^{ih} = e^{ih}\left(e^{i(\ln(1+h)-h)} - 1\right).$$

Note that $ih \le 1$, for all $i = 0 : n$, since $h = \frac{1}{n}$.

Use the Taylor approximations $\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$ and $e^x = 1 + x + O(x^2)$ to obtain that

$$\begin{aligned} i\ (\ln(1+h) - h) &= -i\frac{h^2}{2} + O(h^3); \\ e^{i\ (\ln(1+h)-h)} - 1 &= -i\frac{h^2}{2} + O(h^3). \end{aligned}$$

Conclude that

$$\max_{i=0:n} |e_i| \le \frac{e}{2}\frac{1}{n} + O\left(\frac{1}{n^2}\right),$$

and therefore that

$$\lim_{n\to\infty} \max_{i=0:n} |e_i| = 0.$$

Note that we actually showed that the finite difference discretization is first order convergent, i.e.,

$$\max_{i=0:n} |e_i| = O\left(\frac{1}{n}\right).$$

7. Consider the following second order ODE:

$$\begin{aligned} 3x^2y''(x) - xy'(x) + y(x) &= 0, \quad \forall\, x \in [0,1]; \\ y(0) = 1; \qquad y(1) = \frac{1}{2}. \end{aligned}$$

(i) Partition the interval $[0,1]$ into n equal intervals, corresponding to nodes $x_i = ih$, $i = 0 : n$, where $h = \frac{1}{n}$. Write the finite difference discretization of the ODE at each node x_i, $i = 1 : (n-1)$, using central finite difference approximations for both $y'(x)$ and $y''(x)$.

(ii) If $n = 6$, we find, from the boundary conditions, that $y_0 = 1$ and $y_6 = \frac{1}{2}$. The finite difference discretization scheme presented above will have five equations can be written as a 5×5 linear system $AY = b$. Find A and b.

8. Show that the ODE

$$y''(x) - 2y'(x) + x^2y(x) = 0$$

can be written as

$$Y'(x) = f(x, Y(x)),$$

where

$$Y(x) = \begin{pmatrix} y(x) \\ y'(x) \end{pmatrix} \quad \text{and} \quad f(x, Y(x)) = \begin{pmatrix} 0 & 1 \\ -x^2 & 2 \end{pmatrix} Y(x).$$

9. Consider a six months plain vanilla European call option with strike 18 on a non–dividend–paying underlying asset with spot price 20. Assume that the asset has lognormal distribution with volatility 20% and that interest rates are constant at 5%.

(i) Compute the Greeks of the call option, i.e, Δ, Γ, ρ, vega, and Θ.

Use finite differences to find approximate values for the Greeks. Recall from section 6.3 that

$$\Delta = \frac{\partial C}{\partial S}; \quad \Gamma = \frac{\partial^2 C}{\partial S^2}; \quad \rho = \frac{\partial C}{\partial r}; \quad \text{vega} = \frac{\partial C}{\partial \sigma}; \quad \Theta = -\frac{\partial C}{\partial T};$$

cf. (6.46) for the last formula.

Denote by $C(S, K, T, \sigma, r)$ the value of the call option obtained from the Black–Scholes formula.

(ii) The forward and central difference approximations Δ_f and Δ_c for Δ, and the central difference approximation Γ_c for Γ are

$$\Delta_f = \frac{C(S+dS, K, T, \sigma, r) - C(S, K, T, \sigma, r)}{dS};$$

$$\Delta_c = \frac{C(S+dS, K, T, \sigma, r) - C(S-dS, K, T, \sigma, r)}{2dS};$$

$$\Gamma_c = \frac{C(S+dS, K, T, \sigma, r) - 2C(S, K, T, \sigma, r) + C(S-dS, K, T, \sigma, r)}{(dS)^2}.$$

Compute the approximation errors for the following values of dS:

dS	Δ_f	Δ_c	Γ_c	$\lvert\Delta - \Delta_f\rvert$	$\lvert\Delta - \Delta_c\rvert$	$\lvert\Gamma - \Gamma_c\rvert$
0.1						
0.01						
0.001						
0.0001						
0.00001						
0.000001						

(iv) Let $d\sigma = 0.0001$, $dr = 0.0001$, and $dT = \frac{1}{252}$, i.e., one day. Find the following forward difference approximations for vega, ρ, and Θ:

$$\text{vega}_f = \frac{C(S, K, T, \sigma + d\sigma, r) - C(S, K, T, \sigma, r)}{d\sigma};$$

$$\rho_f = \frac{C(S, K, T, \sigma, r + dr) - C(S, K, T, \sigma, r)}{dr};$$

$$\Theta_f = -\frac{C(S, K, T + dT, \sigma, r) - C(S, K, T, \sigma, r)}{dT}.$$

10. Show that the value of a plain vanilla European call option satisfies the Black–Scholes PDE. In other words, show that

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + (r-q)S\frac{\partial C}{\partial S} - rC = 0,$$

where $C = C(S, t)$ is given by the Black–Scholes formula (3.53), i.e.,

$$C(S, t) = Se^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2),$$

with d_1 and d_2 given by (3.55) and (3.56), respectively.

Hint: Although direct computation can be used to show this result, one could also use the version of the Black–Scholes PDE involving the Greeks, i.e.,

$$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma + (r-q)S\Delta - rV = 0,$$

and substitute the values (3.66), (3.68), and (3.72) for the Greeks.

11. The value at time t of a forward contract struck at K and maturing at time T, on an underlying asset with spot price S paying dividends continuously at the rate q, is

$$f(S, t) = Se^{-q(T-t)} - Ke^{-r(T-t)}.$$

Show that $f(S, t)$ satisfies the Black–Scholes PDE, i.e., show that

$$\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + (r-q)S\frac{\partial f}{\partial S} - rf = 0.$$

Chapter 7

Multivariable calculus: chain rule, integration by substitution, extremum points. Barrier options. Optimality of early exercise.

Chain rule for functions of several variables.

Change of variables for double integrals.

Finding relative extrema for multivariable functions.

7.1 Chain rule for functions of several variables

The One Variable Case

One way of writing the chain rule for functions of one variable is the following: Let $f(x)$ be a differentiable function, and assume that $x = g(t)$, where $g(t)$ is also differentiable. Then $f(x)$ can be regarded as a function of t, i.e., $f(x) = f(g(t))$ and

$$\frac{d}{dt}(f(g(t))) = f'(g(t))\, g'(t),$$

which can also be written as

$$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

Functions of Two Variables

To extend the chain rule to functions of two variables, let $f(x,y)$ be a differentiable function of x and y, and assume that both variables x and y are differentiable functions of another variable t, i.e., $x = g(t)$ and $y = h(t)$, where $g(t)$ and $h(t)$ are differentiable. Then $f(x,y)$ can also be regarded as a function of t, i.e.,

$$f(x,y) = f(g(t),h(t)),$$

The derivative of f with respect to the variable t is given by the following chain rule formula:

$$\frac{d}{dt}(f(g(t),h(t))) = g'(t)\,\frac{\partial f}{\partial x}(g(t),h(t)) + h'(t)\,\frac{\partial f}{\partial y}(g(t),h(t)),$$

which can also be written as

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

Example: Let $f(x,y) = x^2 + y + xy^3$, with $x = e^{2t}$ and $y = t^2$. Compute $\frac{d}{dt}(f(x,y))$.
Answer:

$$\begin{aligned}\frac{d}{dt}(f(x,y)) &= (2x+y^3)\cdot(2e^{2t}) + (1+3xy^2)\cdot(2t)\\ &= 2(2e^{2t}+t^6)\,e^{2t} + 2t(1+3e^{2t}t^4). \quad \Box\end{aligned}$$

Let $f(x,y)$ be a differentiable function. Assume that x and y are differentiable functions of two other variables s and t, i.e.,

$$x = \phi(s,t) \quad \text{and} \quad y = \psi(s,t).$$

Then $f(x,y)$ can be regarded as a function of s and t, i.e.,

$$f(x,y) = f(\phi(s,t),\psi(s,t)).$$

The partial derivatives of f with respect to the variables s and t are given by the following chain rule formulas:

$$\begin{aligned}\frac{\partial f}{\partial s} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \qquad (7.1)\\ &= \frac{\partial f}{\partial x}\frac{\partial \phi}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial \psi}{\partial s}\\ \frac{\partial f}{\partial t} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} \qquad (7.2)\\ &= \frac{\partial f}{\partial x}\frac{\partial \phi}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial \psi}{\partial t}\end{aligned}$$

Functions of n Variables

Let $f(x_1,x_2,\ldots,x_n)$ be a function of n variables. Assume that each variable x_i, $i = 1:n$, is a differentiable function of m other variables denoted by t_1, t_2, ..., t_m. The derivative of the function f with respect to the variable t_j is computed according to the following chain rule formula:

$$\frac{\partial f}{\partial t_j} = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j} \quad \forall\ j = 1:m. \qquad (7.3)$$

Example: Let $f(x_1,x_2,x_3) = x_1^2 + x_1x_2 + x_1x_3 + 2x_3^2$, with

$$x_1(t_1,t_2) = t_1^2 - t_2^2 + 1, \quad x_2(t_1,t_2) = t_2^2 + t_1 + 1, \quad \text{and} \quad x_3(t_1,t_2) = -t_1^2 - 1.$$

Compute $\frac{\partial f}{\partial t_1}$.

Answer: It is easy to see that $\frac{\partial f}{\partial x_1} = 2x_1 + x_2 + x_3$, $\frac{\partial f}{\partial x_2} = x_1$, $\frac{\partial f}{\partial x_3} = x_1 + 4x_3$, and $\frac{\partial x_1}{\partial t_1} = 2t_1$, $\frac{\partial x_2}{\partial t_1} = 1$, $\frac{\partial x_3}{\partial t_1} = -2t_1$. From (7.3), we obtain that

$$\begin{aligned}\frac{\partial f}{\partial t_1} &= \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \frac{\partial f}{\partial x_3}\frac{\partial x_3}{\partial t_1}\\ &= (2x_1+x_2+x_3)\,(2t_1) + x_1 + (x_1+4x_3)\,(-2t_1)\\ &= (t_1^2-t_2^2+t_1+2)\,(2t_1) + (t_1^2-t_2^2+1) + (-3t_1^2-t_2^2-3)\,(-2t_1)\\ &= 8t_1^3 + 3t_1^2 + 10t_1 - t_2^2 + 1.\end{aligned}$$

Alternatively, by direct computation,

$$\begin{aligned}f(t_1,t_2) &= (t_1^2-t_2^2+1)^2 + (t_1^2-t_2^2+1)(t_2^2+t_1+1)\\ &\quad + (t_1^2-t_2^2+1)(-t_1^2-1) + 2(-t_1^2-1)^2\\ &= 2t_1^4 + t_1^3 - t_1t_2^2 + 5t_1^2 - t_2^2 + t_1 + 3.\end{aligned}$$

Then,

$$\frac{\partial f}{\partial t_1} = 8t_1^3 + 3t_1^2 - t_2^2 + 10t_1 + 1. \quad \Box$$

7.2 Change of variables for double integrals

The One Variable Case

Let $f(x)$ be a continuous function, and let $g(s)$ be a continuously differentiable and invertible function[1] mapping an interval $[c,d]$ into the interval $[a,b]$, i.e., $g:[c,d] \to [a,b]$ with

$$s \in [c,d] \longrightarrow x = g(s) \in [a,b].$$

[1]Note that if $g(s)$ is continuous and invertible, then $g(s)$ must be either strictly increasing or strictly decreasing.

The integral of $f(x)$ with respect to the variable x over the interval $[a, b]$ can be written as an integral with respect to the variable s by using the substitution $x = g(s)$ as follows:

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(s))\,g'(s)\,ds; \tag{7.4}$$

see the integration by substitution formula (1.17)[2].

We note that formula (7.4) can also be interpreted as

$$x = g(s), \quad dx = g'(s)\,ds, \quad x = a \longrightarrow s = g^{-1}(a); \quad x = b \longrightarrow s = g^{-1}(b).$$

Functions of Two Variables

Let $f(x, y)$ be a continuous function of two variables to be integrated over a domain $D \subset \mathbb{R}^2$. Let $\Omega \subset \mathbb{R}^2$ be another domain such that there exists an one-to-one and onto function (ϕ, ψ) with continuous first order derivatives mapping Ω into D, i.e., $(\phi, \psi) : \Omega \to D$ with

$$(s, t) \in \Omega \longrightarrow (x, y) = (\phi(s, t), \psi(s, t)) \in D. \tag{7.5}$$

The integral of $f(x, y)$ with respect to the variables x and y over D can be written as an integral with respect to the variables s and t over the domain Ω by using the substitution $(x, y) = (\phi(s, t), \psi(s, t))$ as follows:

$$\int\int_D f(x, y)\,dxdy = \int\int_\Omega h(s, t)\,dsdt. \tag{7.6}$$

To find the appropriate function $h(s, t)$ from (7.6), note that the function f can be written in terms of the variables s and t, i.e.,

$$f(x, y) = f(\phi(s, t), \psi(s, t)).$$

To estimate $dxdy$ in terms of $dsdt$, recall that $(x, y) = (\phi(s, t), \psi(s, t))$, and the gradient of (ϕ, ψ) is the following 2×2 square matrix:

$$D(\phi(s, t), \psi(s, t)) = \begin{pmatrix} \frac{\partial \phi}{\partial s} & \frac{\partial \phi}{\partial t} \\ \frac{\partial \psi}{\partial s} & \frac{\partial \psi}{\partial t} \end{pmatrix}; \tag{7.7}$$

cf. (1.41). Then,

$$dxdy = \left| \frac{\partial \phi}{\partial s}\frac{\partial \psi}{\partial t} - \frac{\partial \phi}{\partial t}\frac{\partial \psi}{\partial s} \right| dsdt. \tag{7.8}$$

[2]If $g(s)$ is increasing, then $g^{-1}(a) = c$ and $g^{-1}(b) = d$. If $g(s)$ is decreasing, then $g^{-1}(a) = d$ and $g^{-1}(b) = c$.

The factor multiplying $dsdt$ in (7.8) is the absolute value of the determinant of the matrix $D(\phi(s, t), \psi(s, t))$ from (7.7), and is called the Jacobian of the mapping $(\phi, \psi) : \Omega \to D$.

Then, the two dimensional change of variables is given by the formula

$$\int\int_D f(x, y)\,dxdy = \int\int_\Omega f(\phi(s, t), \psi(s, t)) \left| \frac{\partial \phi}{\partial s}\frac{\partial \psi}{\partial t} - \frac{\partial \phi}{\partial t}\frac{\partial \psi}{\partial s} \right| dsdt. \tag{7.9}$$

7.2.1 Change of Variables to Polar Coordinates

A commonly used change of variables is the polar coordinates transformation. In two dimensions, given $(x, y) \in \mathbb{R}^2$, we can find $r \in [0, \infty)$ and $\theta \in [0, 2\pi)$ uniquely determined[3] by the following conditions:

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta. \tag{7.10}$$

Using the notation from (7.5), we define $\phi(r, \theta)$ and $\psi(r, \theta)$ as

$$x = \phi(r, \theta) = r\cos\theta, \quad \text{and} \quad y = \psi(r, \theta) = r\sin\theta.$$

From (7.8), we find that the Jacobian of the two dimensional polar coordinates change of variables is

$$\begin{aligned} \left| \frac{\partial \phi}{\partial r}\frac{\partial \psi}{\partial \theta} - \frac{\partial \phi}{\partial \theta}\frac{\partial \psi}{\partial r} \right| &= |\cos\theta \cdot r\cos\theta - (-r\sin\theta) \cdot \sin\theta| \\ &= |r(\cos^2\theta + \sin^2\theta)| = r, \end{aligned}$$

since $\cos^2\theta + \sin^2\theta = 1$ for any Θ.

From (7.9), it follows that

$$\int\int_{\mathbb{R}^2} f(x, y)\,dxdy = \int_0^\infty \int_0^{2\pi} r\,f(r\cos\theta, r\sin\theta)\,d\theta dr. \tag{7.11}$$

Similarly, if we integrate the function $f(x, y)$ over the disk of radius R centered at the origin, i.e., if $(x, y) \in D = D(0, R)$, then the polar coordinates change of variables is $(x, y) = (r\cos\theta, r\sin\theta)$, where $(r, \theta) \in \Omega = [0, R] \times [0, 2\pi)$, and thus

$$\int\int_{D(0,R)} f(x, y)\,dxdy = \int_0^R \int_0^{2\pi} r\,f(r\cos\theta, r\sin\theta)\,d\theta dr. \tag{7.12}$$

Example: Let $D = D(0, 2)$ be the disk of center 0 and radius 2, and let $f(x, y) = 1 - x^2 - y^2$. Compute $\int\int_D f$.

[3]The fact that the point $(0, 0)$ in the (x, y) space is not uniquely mapped in the (r, θ) space is just a technical matter.

Answer: This integral was already computed in section 2.1, using the definition of a double integral. We now use the polar coordinates change of variables $(x,y) = (r\cos\theta, r\sin\theta)$, where $(r,\theta) \in \Omega = [0,2] \times [0,2\pi)$, to compute the integral in a much easier fashion. From (7.12), we obtain that

$$\begin{aligned}\int\int_{D(0,2)} f(x,y)\ dxdy &= \int_0^2 \int_0^{2\pi} r\ (1-(r\cos\theta)^2 - (r\sin\theta)^2)\ d\theta dr \\ &= \int_0^2 \int_0^{2\pi} r(1-r^2)\ d\theta dr \\ &= 2\pi \int_0^2 r(1-r^2)\ dr \\ &= -4\pi. \quad \square\end{aligned}$$

7.3 Finding relative extrema for multivariable functions

The One Variable Case

Let $f : (a,b) \to \mathbb{R}$ be a twice differentiable function such that $f''(x)$ is continuous. A local minimum point for $f(x)$ is a point where $f(x)$ achieves its smallest value over an interval around that point. Similarly, a local maximum point for $f(x)$ is a point where $f(x)$ achieves its largest value over an interval around that point. In mathematical terms,

x_0 is a local minimum for $f(x)$ if and only if $\exists\ \epsilon > 0$ such that

$$f(x_0) \ \le\ f(x), \quad \forall\ x \in (x_0-\epsilon, x_0+\epsilon);$$

x_0 is a local maximum for $f(x)$ if and only if $\exists\ \epsilon > 0$ such that

$$f(x_0) \ \ge\ f(x), \quad \forall\ x \in (x_0-\epsilon, x_0+\epsilon).$$

Any local extremum point (i.e., a local minimum point or a local maximum point) of $f(x)$ must be a critical point for $f(x)$, i.e., a point x_0 where $f'(x_0) = 0$. The question whether a critical point is a local extremum point is answered by the second derivative test:

If $f''(x_0) > 0$, then the critical point x_0 is local minimum point;

If $f''(x_0) < 0$, then the critical point x_0 is local maximum point;

If $f''(x_0) = 0$, anything can happen: the critical point x_0 could be a a local maximum point, a local minimum point, or not a local extremum point.

Example: The point $x_0 = 0$ is a critical point for the following three functions

$$f_1(x) = x^3; \quad f_2(x) = x^4; \quad f_3(x) = -x^4,$$

and has the property that the second derivative of each function at the point $x_0 = 0$ is equal to 0 as well:

$$\begin{aligned} f_1'(0) &= f_1''(0) &= 0; \\ f_2'(0) &= f_2''(0) &= 0; \\ f_3'(0) &= f_3''(0) &= 0. \end{aligned}$$

It is easy to see that
$x_0 = 0$ is neither a local minimum, nor a local maximum point for $f_1(x)$;
$x_0 = 0$ is a local (actually, a global) minimum point for $f_2(x)$;
$x_0 = 0$ is a local (actually, a global) maximum point for $f_3(x)$. $\square$

Functions of Two Variables

Let $U \subset \mathbb{R}^2$ be an open set, e.g., the product of two one dimensional open intervals, or the entire set $\mathbb{R}^2$. Let $f : U \to \mathbb{R}$ be a function with continuous second order partial derivatives. A local minimum point (or a local maximum point) for the function $f(x,y)$ is a point where the function achieves its smallest (or largest) value in a neighborhood of that point.

Definition 7.1. *The point $(x_0,y_0) \in U$ is a local minimum point for the function $f(x,y)$ if and only if there exist $\epsilon > 0$ such that*

$$f(x_0,y_0) \ \le\ f(x,y), \quad \forall\ (x,y) \in (x_0-\epsilon, x_0+\epsilon) \times (y_0-\epsilon, y_0+\epsilon) \subset U.$$

Definition 7.2. *The point (x_0,y_0) is a local maximum point for the function $f(x,y)$ if and only if there exist $\epsilon > 0$ such that*

$$f(x_0,y_0) \ \ge\ f(x,y), \quad \forall\ (x,y) \in (x_0-\epsilon, x_0+\epsilon) \times (y_0-\epsilon, y_0+\epsilon) \subset U.$$

As in the one dimensional case, any local extremum point is a critical point. To define and classify the critical points of $f(x,y)$, recall from (1.39) that the gradient $Df(x,y)$ of $f(x,y)$ is the following 1×2 row vector:

$$Df(x,y) \ =\ \left(\frac{\partial f}{\partial x}(x,y) \quad \frac{\partial f}{\partial y}(x,y)\right),$$

and, from (1.40), that the Hessian $D^2 f(x,y)$ of $f(x,y)$ is the 2×2 square matrix given by

$$D^2 f(x,y) \ =\ \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial y \partial x}(x,y) \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix}.$$

Definition 7.3. *The point (x_0, y_0) is a critical point for the function $f(x,y)$ if and only if*

$$Df(x_0, y_0) \;=\; 0,$$

which is equivalent to

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Lemma 7.1. *Every local extremum point for the function $f(x,y)$ is a critical point.*

Critical points are classified using the following two dimensional version of the second derivative test:

Theorem 7.1. *Let (x_0, y_0) be a critical point for the function $f(x,y)$ and assume that all second order partial derivatives of $f(x,y)$ exist and are continuous.*

If the matrix $D^2 f(x_0, y_0)$ is positive definite, i.e., if both eigenvalues of the matrix $D^2 f(x_0, y_0)$ are strictly positive, then the critical point x_0 is a local minimum point;

If the matrix $D^2 f(x_0, y_0)$ is negative definite, i.e., if both eigenvalues of the matrix $D^2 f(x_0, y_0)$ are strictly negative, then the critical point x_0 is a local maximum point;

If the two eigenvalues of the matrix $D^2 f(x_0, y_0)$ are nonzero and of opposite signs, then the critical point (x_0, y_0) is a saddle point, and it is not an extremum point;

If one eigenvalues of the matrix $D^2 f(x_0, y_0)$ is equal to 0 and the other eigenvalue is strictly positive, then the critical point x_0 is local minimum point. If the other eigenvalue is strictly negative, then the critical point x_0 is local maximum point. If both eigenvalues are 0, i.e., if the matrix $D^2 f(x_0, y_0)$ is itself equal to 0, anything could happen, i.e., the critical point (x_0, y_0) could be a local maximum point, a local minimum point, or a saddle point.

While the eigenvalues of a matrix with real entries are, in general, complex numbers, the eigenvalues of a symmetric matrix are real numbers. Note that $D^2 f(x_0, y_0)$ is a symmetric matrix, i.e., $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$, since $\frac{\partial^2 f}{\partial y \partial x}$ and $\frac{\partial^2 f}{\partial x \partial y}$ are continuous; cf. Theorem 1.9. Therefore, the eigenvalues of $D^2 f(x_0, y_0)$ are real numbers and it makes sense to talk about their signs.

The results of Theorem 7.1 are often stated in terms of $\det(D^2 f(x_0, y_0))$, the determinant of the Hessian of f evaluated at (x_0, y_0).

Theorem 7.2. *Let (x_0, y_0) be a critical point for the function $f(x,y)$ and assume that all second order derivatives of $f(x,y)$ exist and are continuous.*

If $\det(D^2 f(x_0, y_0)) > 0$ and $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$, then the critical point (x_0, y_0) is a local minimum point;

If $\det(D^2 f(x_0, y_0)) > 0$ and $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0$, then the critical point (x_0, y_0) is a local maximum point;

If $\det(D^2 f(x_0, y_0)) < 0$, the critical point (x_0, y_0) is called a saddle point, and it is not an extremum point;

If $\det(D^2 f(x_0, y_0)) = 0$, anything could happen, i.e., the critical point (x_0, y_0) could be a a local maximum point, a local minimum point, or a saddle point.

We note that Theorem 7.2 can be regarded as a corollary of Theorem 7.1. This can be seen from the following result:

Lemma 7.2. *Let A be a 2×2 symmetric matrix as follows:*

$$A \;=\; \begin{pmatrix} a & b \\ b & d \end{pmatrix}.$$

The eigenvalues of A have the same sign (and are nonzero) if and only if the determinant of A is strictly positive.

If $\det(A) > 0$ and $a > 0$, then both eigenvalues of A are strictly positive.

If $\det(A) > 0$ and $a < 0$, then both eigenvalues of A are strictly negative.

Proof. Let λ_1 and λ_2 be the eigenvalues of A. We do not assume that λ_1 and λ_2 are necessarily different. The determinant of any matrix is equal to the product of its eigenvalues, i.e.,

$$\det(A) \;=\; \lambda_1 \, \lambda_2.$$

Therefore, λ_1 and λ_2 have the same sign and are nonzero if and only if $\det(A)$ is strictly positive, i.e., $\det(A) > 0$.

We recall that, by definition,

$$\det(A) \;=\; ad \;-\; b^2.$$

Then, if $\det(A) > 0$, we find that $a \neq 0$, since otherwise $\det(A) = -b^2 < 0$.

Proving that if $\det(A) > 0$ and $a > 0$ then both eigenvalues of A are positive is more subtle. A short and insightful proof requires knowledge of the following result:

There exists an upper triangular matrix U with all entries on the main diagonal strictly positive such that $A = U^t U$ if and only if all the eigenvalues of the matrix A are strictly positive[4].

[4]The decomposition $U^t U = A$ is called the Cholesky decomposition of the matrix A.

Assume that $a > 0$ and $\det(A) = ad - b^2 > 0$. Then, the matrix

$$U = \begin{pmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \frac{\sqrt{ad-b^2}}{\sqrt{a}} \end{pmatrix}$$

is upper triangular and has all entries on the main diagonal strictly positive. By matrix multiplication, we obtain that

$$U^t\, U = \begin{pmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \frac{\sqrt{ad-b^2}}{\sqrt{a}} \end{pmatrix} \begin{pmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \frac{\sqrt{ad-b^2}}{\sqrt{a}} \end{pmatrix} = A,$$

and conclude that both eigenvalues of A are strictly positive in this case.

To complete the proof, we need to show that, if $a < 0$ and $\det(A) = ad - b^2 > 0$, then both eigenvalues of A are strictly negative.

Let $A_1 = -A$. Then $\det(A_1) = \det(A)$ and the upper leftmost entry of the matrix A_1 is $-a > 0$. From the result proved above, we find that both eigenvalues of A_1 are strictly positive. Therefore, the eigenvalues of A, which are the same in absolute value as the eigenvalues of A_1, but have opposite signs, are strictly negative. □

Example: The point $(0,0)$ is a critical point for the following functions:

$$\begin{aligned} f_1(x) &= x^2 + xy + y^2 \\ f_2(x) &= -x^2 - xy - y^2 \\ f_3(x) &= x^2 + 3xy + y^2 \end{aligned}$$

The Hessians of these functions (which are constant matrices) evaluated at the critical point $(0,0)$, and their determinants, are as follows:

$$D^2 f_1(0,0) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}; \quad \det(D^2 f_1(0,0)) = 3;$$

$$D^2 f_2(0,0) = \begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix}; \quad \det(D^2 f_2(0,0)) = 3;$$

$$D^2 f_3(0,0) = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}; \quad \det(D^2 f_3(0,0)) = -5.$$

From Theorem 7.2, it follows that the point $(0,0)$ is a local minimum for $f_1(x,y)$, a local maximum for $f_2(x,y)$, and a saddle point for $f_3(x,y)$. □

While we do not give formal proofs for Lemma 7.1 and Theorem 7.1, we will provide the intuition behind these results.

Recall the matrix form of the quadratic Taylor expansion (5.41) of a function of two variables. By writing this expansion for the function $f(x,y)$ around the point $(a,b) = (x_0, y_0)$, we find that

$$\begin{aligned} f(x,y) - f(x_0,y_0) &= Df(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &\quad + \frac{1}{2} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}^t D^2 f(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &\quad + O\left(|x - x_0|^3\right) + O\left(|y - y_0|^3\right), \end{aligned} \tag{7.13}$$

where $Df(x_0,y_0)$ and $D^2 f(x_0,y_0)$ are the gradient and Hessian of $f(x,y)$ evaluated at the point (x_0,y_0), respectively.

Assume that $(x_0,y_0) \in U$ is a local extremum point for $f(x,y)$, e.g., a local minimum point. Then, by definition, there exists an interval I around (x_0,y_0) such that $f(x,y) \geq f(x_0,y_0)$ for all points $(x,y) \in I$. From (7.13), we find that, for all $(x,y) \in I$,

$$\begin{aligned} Df(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} &+ \frac{1}{2} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}^t D^2 f(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &+ O\left(|x - x_0|^3\right) + O\left(|y - y_0|^3\right) \geq 0. \end{aligned} \tag{7.14}$$

If we let $(x,y) \in I$ be close to the point (x_0,y_0), i.e., if $|x - x_0|$ and $|y - y_0|$ are very small, then the dominant term in (7.14) is the linear term

$$Df(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \tag{7.15}$$

If $Df(x_0,y_0) \neq 0$, the inequality (7.14) cannot hold, since we can always choose x and y close to x_0 and y_0, respectively, such that the dominant (and nonzero) term (7.15) has negative sign. The entire left hand side of (7.14) would therefore be negative, which contradicts (7.14), and therefore the point (x_0,y_0) cannot be a local minimum point.

We conclude that, if (x_0,y_0) is a local minimum point for $f(x,y)$, then $Df(x_0,y_0) = 0$, and therefore (x_0,y_0) is a critical point for the function $f(x,y)$, as stated by Lemma 7.1. The intuition for the local maximum case follows along the exact same lines of thought.

We showed that, if (x_0,y_0) is a critical point for $f(x,y)$, then $Df(x_0,y_0) = 0$. Formula (7.13) becomes

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + \frac{1}{2} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}^t D^2 f(x_0,y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &\quad + O\left(|x - x_0|^3\right) + O\left(|y - y_0|^3\right). \end{aligned} \tag{7.16}$$

We now show that, if the matrix $D^2f(x_0,y_0)$ is positive definite, then the critical point (x_0,y_0) is a local minimum point. If $|x-x_0|$ and $|y-y_0|$ are very small, then the terms $O\left(|x-x_0|^3\right)$ and $O\left(|y-y_0|^3\right)$ from (7.16) are dominated by the term

$$\frac{1}{2}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}^t D^2f(x_0,y_0)\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} \;>\; 0; \qquad (7.17)$$

the inequality from (7.17) follows from the fact that $D^2f(x_0,y_0)$ is a positive definite matrix[5]. From (7.16), and using (7.17), we can conclude that, for any point (x,y) close to the point (x_0,y_0),

$$f(x,y)-f(x_0,y_0) \;\approx\; \frac{1}{2}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}^t D^2f(x_0,y_0)\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} \;>\; 0.$$

Therefore, the point (x_0,y_0) is a local minimum point for the function $f(x,y)$.

The intuition for all the other cases from Theorem 7.1 follow from a similar analysis of the Taylor expansion (7.16).

Example: Find and classify the local extrema of the function $f:\mathbb{R}^2\to\mathbb{R}$ given by

$$f(x,y) \;=\; x^2+2xy+5y^2+2x+10y-5.$$

Answer: The function $f(x,y)$ has only one critical point, $(0,-1)$, which is the unique solution of the linear system

$$\begin{cases} \frac{\partial f}{\partial x} \;=\; 2x+2y+2 \;=\; 0; \\ \frac{\partial f}{\partial y} \;=\; 2x+10y+10 \;=\; 0. \end{cases}$$

Note that $D^2f(0,-1)=\begin{pmatrix} 2 & 2 \\ 2 & 10 \end{pmatrix}$ and $\det(D^2f(0,-1))=16>0$. Since the upper leftmost entry of the Hessian $D^2f(0,-1)$ is equal to $2>0$, we conclude from Theorem 7.2 that the point $(0,-1)$ is a local minimum point for the function $f(x,y)$. $\square$

Functions of n variables

The results for extrema of functions of two variables obtained above extend naturally to the case of functions of n variables.

Let $U\subset\mathbb{R}^n$ be an open set[6]. Let $f:U\to\mathbb{R}$ be a twice differentiable function, with continuous second order partial derivatives.

[5]Recall that A is a positive definite matrix if and only if $v^tAv>0$, for any $v\neq 0$.

[6]The proper definition of an open set $U\subset\mathbb{R}^n$ is as follows: For any $x\in U$, there exists an interval I around x such that the entire interval I is contained in U. Examples of open sets are n-dimensional open intervals, i.e., $U=\prod_{i=1:n}(a_i,b_i)$, and the set of n-dimensional vectors with positive entries, i.e., $U=\prod_{i=1:n}(0,\infty)$. The space $\mathbb{R}^n$ is also an open set.

A point $x_0\in U$ is a local minimum for the function $f(x)$ if and only if there exists a neighborhood $I\subset U$ of x_0 such that $f(x_0)\leq f(x)$, for all $x\in I$.

A point $x_0\in U$ is a local maximum for the function $f(x)$ if and only if there exists a neighborhood I of x_0 such that $f(x_0)\geq f(x)$, for all $x\in I$.

Recall that the gradient $Df(x)$ and Hessian $D^2f(x)$ of the function $f(x)$ are given by (1.36) and (1.37), respectively.

Lemma 7.3. *Every local extremum point x_0 for the function $f(x)$ must be a critical point, i.e., $Df(x_0)=0$.*

We recall that if A is symmetric and has real entries, then all its the eigenvalues of A are real numbers. (In general, an $n\times n$ matrix with real entries has n eigenvalues, counted with their multiplies, but these eigenvalues may very well be complex numbers.)

The second derivative test for n–dimensional functions is given below:

Theorem 7.3. *Let $f:U\to\mathbb{R}$ be a twice differentiable function, with continuous second order partial derivatives, and let x_0 be a critical point for the function $f(x)$. Then, the Hessian matrix $D^2f(x_0)$ of f evaluated at x_0, is a symmetric matrix.*

If all eigenvalues of $D^2f(x_0)$ are greater than or equal to 0, and at least one in nonzero, then x_0 is a local minimum point.

If all eigenvalues of $D^2f(x_0)$ are smaller than or equal to 0, and at least one in nonzero, then x_0 is a local maximum point.

If the matrix $D^2f(x_0)$ has both positive and negative eigenvalues, then the point x_0 is a saddle point, i.e., neither a local minimum, nor a local maximum.

FINANCIAL APPLICATIONS

The Theta of a derivative security.

Show that the density function of the standard normal variable has unit integral over $\mathbb{R}$.

The Box–Muller method for generating samples of standard normal variables.

Reducing the Black–Scholes PDE to the heat equation.

Barrier options. Arbitrage arguments for the up-and-in up-and-out parity.

On the optimality of early exercise for American Call options.

7.4 The Theta of a derivative security

The Theta of a derivative security is the rate of change of the value V of the security with respect to time t, i.e.,

$$\Theta(V) = \frac{\partial V}{\partial t}. \tag{7.18}$$

Recall from (3.72) that the Theta of a plain vanilla European call option is

$$\Theta(C) = -\frac{S\sigma e^{-q(T-t)}}{2\sqrt{2\pi(T-t)}}e^{-\frac{d_1^2}{2}} + qSe^{-q(T-t)}N(d_1) - rKe^{-r(T-t)}N(d_2), \tag{7.19}$$

where d_1 and d_2 given by (3.55) and (3.56), respectively, i.e.,

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r - q + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}. \tag{7.20}$$

Of all the Greeks, Theta causes most confusions, mainly generated by using a simplified version of the Black–Scholes formula when computing Theta. The Black–Scholes formula (3.57) is often written for time $t = 0$, i.e.,

$$C = Se^{-qT}N(d_1) - Ke^{-rT}N(d_2), \tag{7.21}$$

where d_1 and d_2 are given by (7.20) for time $t = 0$, i.e.,

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r - q + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}.$$

The Theta of the call option is defined as the rate of change of C with respect to the passage of time and given as

$$\Theta(C) = -\frac{S\sigma e^{-qT}}{2\sqrt{2\pi T}}e^{-\frac{d_1^2}{2}} + qSe^{-qT}N(d_1) - rKe^{-rT}N(d_2); \tag{7.22}$$

this formula is correct, and corresponds to (7.19) with $t = 0$.

We note that, since $\Theta(C)$ is defined as

$$\Theta(C) = \frac{\partial C}{\partial t},$$

this would imply that (7.22) should be obtained by differentiating the formula (7.21) with respect to t. However, this would give $\Theta(C) = 0$, since (7.21) does not depend on the variable t.

The explanation is that (7.21) should be regarded as the formula for $C(S,t)$ evaluated at time $t = 0$, where $C(S,t)$ is given by the Black–Scholes formula (3.53), i.e.,

$$C(S,t) = Se^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2), \tag{7.23}$$

with d_1 and d_2 given by (7.20). Then, $\Theta(C)$ at time $t = 0$ is given by the partial derivative of $C(S,t)$ from (7.23) with respect to t, evaluated at $t = 0$, i.e.,

$$\Theta(C) = \left.\frac{\partial C(S,t)}{\partial t}\right|_{t=0}.$$

This yields formula (7.19) for $\Theta(C)$ evaluated at $t = 0$, which is the same as (7.22).

Another point that requires attention is that $\Theta(C)$ is sometimes given as

$$\Theta(C) = -\frac{\partial C}{\partial T}. \tag{7.24}$$

This formula is mathematically correct, as it can be seen by differentiating (7.21) and (7.23) with respect to T to obtain (7.22) and (7.19), respectively. While formula (7.24) also holds for the case $t = 0$, it would be confusing to interpret it as Theta being the derivative of the price of the option with respect to T, the maturity of the option.

The correct interpretation is that $\Theta(C)$ is the rate of change of the value of the call option with respect to time left to maturity $T - t$. Mathematically, this corresponds to

$$\Theta(C) = -\frac{\partial C}{\partial(T-t)}. \tag{7.25}$$

By differentiating formula (7.23) with respect to the variable $T-t$, it is easy to see that (7.25) yields the formula (7.19) for $\Theta(C)$. A direct and more insightful proof of the fact that (7.25) is correct can be given using chain rule as follows:

$$\frac{\partial C(S,t)}{\partial t} = \frac{\partial C(S,t)}{\partial(T-t)} \cdot \frac{\partial(T-t)}{\partial t} = -\frac{\partial C(S,t)}{\partial(T-t)}.$$

Here, T is a fixed constant and t is considered to be a variable.

As for the mathematical correctness of formula (7.24), note that, in (7.24), C is regarded as a function of T, and not as a function of t, as was the case in (7.23). Then, considering t a fixed constant, we apply chain rule and find that

$$\frac{\partial C}{\partial T} = \frac{\partial C}{\partial(T-t)}\,\frac{\partial(T-t)}{\partial T} = \frac{\partial C}{\partial(T-t)}. \tag{7.26}$$

Using (7.26) and the definition (7.25) for $\Theta(C)$, we conclude that formula (7.24) is mathematically correct:

$$\Theta(C) = -\frac{\partial C}{\partial(T-t)} = -\frac{\partial C}{\partial T}.$$

We emphasize again that while formula (7.24) holds true, the financially insightful formula to use is (7.25), which shows that the Theta of the option is equal to the negative rate of change of the value of the option with respect to the time left until maturity.

7.5 Integrating the density function of the standard normal variable

Recall from section 3.3 that the probability density function of the standard normal variable is

$$f(t) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}.$$

We want to show that $f(t)$ is indeed a density function. It is clear that $f(t) \geq 0$ for any $t \in \mathbb{R}$. According to (3.33), we also have to prove that

$$\int_{-\infty}^{\infty} f(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = 1. \tag{7.27}$$

We use the substitution $t = \sqrt{2}\,x$. Then $dt = \sqrt{2}dx$, and $t = -\infty$ and $t = \infty$ are mapped into $x = -\infty$ and $x = \infty$, respectively. Therefore,

$$\int_{-\infty}^{\infty} f(t)\, dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2}\, dx.$$

Thus, in order to prove (7.27), we only need to show that

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}. \tag{7.28}$$

Rather surprisingly, the polar coordinates change of variables can be used to prove (7.28). Let

$$I = \int_{-\infty}^{\infty} e^{-x^2}\, dx.$$

We want to show that

$$I = \sqrt{\pi}.$$

Since x is just an integrating variable, we can also write the integral I in terms of another integrating variable, denoted by y, as follows:

$$I = \int_{-\infty}^{\infty} e^{-y^2}\, dy.$$

Then,

$$\begin{aligned}
I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2}\, dx\right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2}\, dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2} e^{-y^2}\, dxdy \\
&= \int\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\, dxdy.
\end{aligned}$$

(Note that, while the second equality may be intuitively clear, a result similar to Theorem 2.1 is required for a rigorous derivation.)

We use the polar coordinates transformation (7.10) to evaluate the last integral. We change the variables $(x, y) \in \mathbb{R}^2$ to $(r, \theta) \in [0,\ \infty) \times [0,\ 2\pi)$ given by $x = r\cos\theta$ and $y = r\sin\theta$. From (7.11), we find that

$$\begin{aligned}
I^2 &= \int\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\, dxdy \\
&= \int_0^{\infty}\int_0^{2\pi} r\, e^{-\left((r\cos\theta)^2+(r\sin\theta)^2\right)}\, d\theta dr \\
&= \int_0^{\infty}\int_0^{2\pi} r\, e^{-\left(r^2\,(\cos^2\theta+\sin^2\theta)\right)}\, d\theta dr \\
&= \int_0^{\infty}\int_0^{2\pi} r\, e^{-r^2}\, d\theta dr \\
&= \int_0^{\infty} 2\pi\, r\, e^{-r^2}\, dr \\
&= 2\pi \lim_{t\to\infty} \int_0^t r\, e^{-r^2}\, dr \\
&= 2\pi \lim_{t\to\infty} \left(-\frac{1}{2}e^{-r^2}\right)\Bigg|_0^t \\
&= \pi.
\end{aligned}$$

Thus, we proved that $I^2 = \pi$ and therefore that $I = \sqrt{\pi}$, i.e.,

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = I = \sqrt{\pi}.$$

As shown above, this is equivalent to showing that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt \;=\; 1,$$

which completes the proof of the fact that $f(t) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}$ is a density function.

7.6 The Box–Muller method

The Box–Muller method is an algorithm for generating samples of the standard normal variable, which are needed for Monte Carlo simulations. In the Box–Muller method, two independent standard normal samples are generated, starting with two independent samples from the uniform distribution. The method is based on the following fact:

If Z_1 and Z_2 are independent standard normal variables, then $R = Z_1^2 + Z_2^2$ is an exponential random variable with mean 2, and, given R, the point (Z_1, Z_2) is uniformly distributed on the circle of center 0 and radius $\sqrt{R}$.

Explaining the Box–Muller method and the Marsaglia polar method for an efficient implementation of the Box–Muller algorithm is beyond the scope of this book. We resume the discussion to applying the polar coordinates change of variable to explain the following part of the Box–Muller method:

Lemma 7.4. *Let Z_1 and Z_2 be independent standard normal variables, and let $R = Z_1^2 + Z_2^2$. Then R is an exponential random variable with mean 2.*

Proof. Recall that the cumulative density function of an exponential random variable X with parameter $\alpha > 0$ is

$$F(x) \;=\; \begin{cases} 1 - e^{-\alpha x}, & \text{if } \; x \;\geq\; 0; \\ 0, & \text{otherwise,} \end{cases}$$

and its expected value is $E[X] = \frac{1}{\alpha}$.

Let $R = Z_1^2 + Z_2^2$. To prove that R is an exponential random variable with mean $E[R] = 2$, i.e., with parameter $\alpha = \frac{1}{2}$, we will show that the cumulative density of R is

$$F(x) \;=\; P(R \leq x) \;=\; \begin{cases} 1 - e^{-\frac{x}{2}}, & \text{if } \; x \;\geq\; 0; \\ 0, & \text{otherwise.} \end{cases} \tag{7.29}$$

Let $x < 0$. Since $R = Z_1^2 + Z_2^2 \geq 0$, we find that $P(R \leq x) = 0$.

Let $x \geq 0$. Let $f_1(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right)$ and $f_2(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right)$ be the density functions of Z_1 and Z_2, respectively, where $\exp(t) = e^t$. Since Z_1 and Z_2 are independent, the joint density function of Z_1 and Z_2 is the product function $f_1(x_1)f_2(x_2)$; cf. Lemma 4.5. We note that $Z_1^2 + Z_2^2 \leq x$ if and only if the point (Z_1, Z_2) is in the disk $D(0, \sqrt{x})$ of center 0 and radius $\sqrt{x}$, i.e., if $(Z_1, Z_2) \in D(0, \sqrt{x})$. Then,

$$\begin{aligned} P(R \leq x) \;&=\; P(Z_1^2 + Z_2^2 \;\leq\; x) \\ &=\; \int\!\!\int_{D(0,\sqrt{x})} f_1(x_1) f_2(x_2)\; dx_1\; dx_2 \\ &=\; \frac{1}{2\pi} \int\!\!\int_{D(0,\sqrt{x})} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)\; dx_1 dx_2. \end{aligned}$$

The polar coordinates change of variables $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$ maps any point (x_1, x_2) from the disk $D(0, \sqrt{x})$ into a point (r, θ) from the domain $[0,\ \sqrt{x}] \times [0,\ 2\pi)$. From (7.12), we find that

$$\begin{aligned} P(R \leq x) \;&=\; \frac{1}{2\pi} \int\!\!\int_{D(0,\sqrt{x})} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)\; dx_1 dx_2 \\ &=\; \frac{1}{2\pi} \int_0^{\sqrt{x}} \int_0^{2\pi} r\, \exp\left(-\frac{(r\cos\theta)^2 + (r\sin\theta)^2}{2}\right)\; d\theta dr \\ &=\; \frac{1}{2\pi} \int_0^{\sqrt{x}} \int_0^{2\pi} r\, \exp\left(-\frac{r^2}{2}\right)\; d\theta dr \\ &=\; \frac{1}{2\pi}\, 2\pi \int_0^{\sqrt{x}} r\; e^{-\frac{r^2}{2}}\; dr \\ &=\; \left(-e^{-\frac{r^2}{2}}\right)\Big|_0^{\sqrt{x}} \\ &=\; 1 \;-\; e^{-\frac{x}{2}}. \end{aligned}$$

We proved that the cumulative density of the random variable R is equal to $1 - e^{-\frac{x}{2}}$, if $x \geq 0$, and to 0 otherwise. This is equivalent to showing that $R = Z_1^2 + Z_2^2$ is an exponential random variable with mean 2; cf. (7.29). □

7.7 Reducing the Black–Scholes PDE to the heat equation

Let $V(S, t)$ be the value at time t of a European call or put option with maturity T, on a lognormally distributed underlying asset with spot price

S and volatility σ, paying dividends continuously at rate q. Recall from section 6.4 that $V(S,t)$ satisfies the Black–Scholes PDE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV = 0, \quad \forall\, S > 0, \quad \forall\, 0 < t < T. \tag{7.30}$$

The risk-free interest rate r is assumed to be constant over the lifetime of the option. The boundary condition at maturity is $V(S,T) = \max(S-K,0)$, for the call option and $V(S,T) = \max(K-S,0)$, for the put option.

One way to solve the Black–Scholes PDE (7.30) is to use a change of variables to reduce it to a boundary value problem for the heat equation

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}, \tag{7.31}$$

since a closed form solution to this problem is known.

The change of variables is as follows:

$$V(S,t) = \exp(-ax - b\tau)u(x,\tau), \tag{7.32}$$

where

$$x = \ln\left(\frac{S}{K}\right), \tag{7.33}$$

$$\tau = \frac{(T-t)\sigma^2}{2}, \tag{7.34}$$

and the constants a and b are given by

$$a = \frac{r-q}{\sigma^2} - \frac{1}{2}; \tag{7.35}$$

$$b = \left(\frac{r-q}{\sigma^2} + \frac{1}{2}\right)^2. \tag{7.36}$$

A brief explanation of how the change of variables (7.32) is chosen is in order.

In general, changing from the variables space (S,t) to the space (x,τ) means choosing functions $\phi(S,t)$ and $\psi(S,t)$ such that $x = \phi(S,t)$ and $\tau = \psi(S,t)$. A simpler version of such change of variables is to have x and τ depend on only one variable each, i.e., $x = \phi(S)$ and $\tau = \psi(t)$.

Note that the Black–Scholes PDE (7.30) has nonconstant coefficients, which makes it more challenging to solve. However, the Black–Scholes PDE is homogeneous, in the sense that all the terms with nonconstant coefficients are of the form

$$S^2\frac{\partial^2 V}{\partial S^2} \quad \text{and} \quad S\frac{\partial V}{\partial S}.$$

The classical change of variable for ODEs with homogeneous nonconstant coefficients involves the logarithmic function. In our case, we choose the change of variables (7.33), i.e.,

$$x = \ln\left(\frac{S}{K}\right).$$

From a mathematical standpoint, the change of variables (7.33) is equivalent to $x = \ln(S)$. The term $\frac{S}{K}$ is introduced for financial reasons: taking the logarithm of a dollar amount, as would be the case for $\ln(S)$, does not make sense. However, $\frac{S}{K}$ is a non–denomination quantity, i.e., a number, and its logarithm, $\ln\left(\frac{S}{K}\right)$, is well defined.

The Black–Scholes PDE (7.30) is backward in time, i.e., the boundary data is given at time T and the solution is required at time 0. We choose the change of variables for τ to obtain a forward PDE in τ, i.e., with boundary data given at $\tau = 0$ and solution required a fixed time $\tau_{final} > 0$. The change of variables (7.34), i.e.,

$$\tau = \frac{(T-t)\sigma^2}{2},$$

accomplishes this, and will also cancel out the term $\frac{\sigma^2}{2}$ from the coefficient of the second order partial derivative.

We now prove that, if $V(S,t)$ satisfies the Black–Scholes PDE (7.30), then the function $u(x,\tau)$ given by (7.32) satisfies the heat equation (7.31).

We do this in two stages. First, we make the change of variables

$$V(S,t) = w(x,\tau),$$

where

$$x = \ln\left(\frac{S}{K}\right) \quad \text{and} \quad \tau = \frac{(T-t)\sigma^2}{2}.$$

It is easy to see that

$$\frac{\partial x}{\partial S} = \frac{1}{S}; \quad \frac{\partial x}{\partial t} = 0; \tag{7.37}$$

$$\frac{\partial \tau}{\partial S} = 0; \quad \frac{\partial \tau}{\partial t} = -\frac{\sigma^2}{2} \tag{7.38}$$

Using chain rule repeatedly, see (7.1) and (7.2), as well as (7.37) and (7.38) we find that

$$\frac{\partial V}{\partial S} = \frac{\partial w}{\partial S} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial S} + \frac{\partial w}{\partial \tau}\frac{\partial \tau}{\partial S} = \frac{1}{S}\frac{\partial w}{\partial x}$$

$$\frac{\partial^2 V}{\partial S^2} = \frac{\partial}{\partial S}\left(\frac{\partial V}{\partial S}\right) = \frac{\partial}{\partial S}\left(\frac{1}{S}\frac{\partial w}{\partial x}\right)$$

$$\begin{aligned}
&= -\frac{1}{S^2}\frac{\partial w}{\partial x} + \frac{1}{S}\frac{\partial}{\partial S}\left(\frac{\partial w}{\partial x}\right)\\
&= -\frac{1}{S^2}\frac{\partial w}{\partial x} + \frac{1}{S}\left(\frac{\partial^2 w}{\partial x^2}\frac{\partial x}{\partial S} + \frac{\partial^2 w}{\partial \tau\, \partial x}\frac{\partial \tau}{\partial S}\right)\\
&= -\frac{1}{S^2}\frac{\partial w}{\partial x} + \frac{1}{S^2}\frac{\partial^2 w}{\partial x^2}\\
&= \frac{1}{S^2}\left(\frac{\partial^2 w}{\partial x^2} - \frac{\partial w}{\partial x}\right)
\end{aligned}$$

$$\frac{\partial V}{\partial t} = \frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial \tau}\frac{\partial \tau}{\partial t} = -\frac{\sigma^2}{2}\frac{\partial w}{\partial \tau}$$

By substituting in (7.30), we obtain the following PDE satisfied by $w(x, \tau)$:

$$-\frac{\sigma^2}{2}\frac{\partial w}{\partial \tau} + \frac{\sigma^2}{2}S^2\cdot\frac{1}{S^2}\left(\frac{\partial^2 w}{\partial x^2} - \frac{\partial w}{\partial x}\right) + (r-q)S\cdot\frac{1}{S}\frac{\partial w}{\partial x} - rw = 0. \quad (7.39)$$

After canceling out the terms involving S and dividing by $-\frac{\sigma^2}{2}$, the PDE (7.39) becomes

$$\frac{\partial w}{\partial \tau} - \frac{\partial^2 w}{\partial x^2} + \left(1 - \frac{2(r-q)}{\sigma^2}\right)\frac{\partial w}{\partial x} + \frac{2r}{\sigma^2}w = 0. \quad (7.40)$$

The PDE (7.40) satisfied by the function $w(x, \tau)$ has constant coefficients and is forward parabolic. To eliminate the lower order terms, i.e., the terms corresponding to $\frac{\partial w}{\partial x}$ and w, let

$$w(x, \tau) = \exp(-ax - b\tau)u(x, \tau),$$

where a and b are constants to be determined later.

It is easy to see that

$$\begin{aligned}
\frac{\partial w}{\partial x} &= \exp(-ax - b\tau)\left(-au + \frac{\partial u}{\partial x}\right);\\
\frac{\partial^2 w}{\partial x^2} &= \exp(-ax - b\tau)\left(a^2 u - 2a\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2}\right);\\
\frac{\partial w}{\partial \tau} &= \exp(-ax - b\tau)\left(-bu + \frac{\partial u}{\partial \tau}\right).
\end{aligned}$$

Therefore, the PDE (7.40) for $w(x, \tau)$ becomes the following PDE for $u(x, \tau)$:

$$\begin{aligned}
&\frac{\partial u}{\partial \tau} - \frac{\partial^2 u}{\partial x^2} + \left(2a + 1 - \frac{2(r-q)}{\sigma^2}\right)\frac{\partial u}{\partial x}\\
&- \left(b + a^2 + a\left(1 - \frac{2(r-q)}{\sigma^2}\right) - \frac{2r}{\sigma^2}\right)u = 0.
\end{aligned}$$

We choose the constants a and b is such a way that the coefficients of $\frac{\partial u}{\partial x}$ and u are equal to 0, i.e., such that

$$\begin{cases} 2a + 1 - \frac{2(r-q)}{\sigma^2} = 0\\ b + a^2 + a\left(1 - \frac{2(r-q)}{\sigma^2}\right) - \frac{2r}{\sigma^2} = 0 \end{cases}$$

The solution of this system is given by (7.35–7.36); see an exercise at the end of this chapter. Thus, using the change of variables (7.32), the Black–Scholes PDE (7.30) for $V(S, t)$ becomes the heat equation for $u(x, \tau)$, i.e.,

$$\frac{\partial u}{\partial \tau} - \frac{\partial^2 u}{\partial x^2} = 0.$$

While this is all we wanted to show here, we note that we can use the same change of variables to transform the boundary conditions $V(S, T)$ into boundary conditions $u(x, 0)$ for the heat equation. A closed formula for solving the heat equation with boundary conditions at time $t = 0$ exists. Therefore, closed formulas for $u(x, \tau)$ can be obtained, and, from (7.32), closed formulas for $V(S, t)$ can then be inferred. If the boundary conditions for $V(S, T)$ are chosen to correspond to those for European call or put options payoffs at time T, then the Black–Scholes formulas (3.57) and (3.58), previously derived in section 4.7 by using risk–neutral pricing, are obtained.

7.8 Barrier options

An exotic option is any option that is not a plain vanilla option. Some exotic options are path–dependent: the value of the option depends not only on the value of the underlying asset at maturity, but also on the path followed by the price of the asset between the inception of the option and maturity.

It rarely happens to have closed formulas for pricing exotic options. However, European barrier options are path–dependent options for which closed form pricing formulas exist. A barrier option is different from a plain vanilla option due to the existence of a *barrier* B: the option either expires worthless (**knock–out options**), or becomes a plain vanilla option (**knock–in options**) if the price of the underlying asset hits the barrier before maturity. Depending on whether the barrier must be hit from below or from above in order to be triggered, an option is called **up** or **down**. Every option can be either a call or a put option, so there are, at the first count, at least eight different types of options:

up *in* *put*
down *out* *call*

This should be read as follows: choose one entry in each column, e.g., up, out, call. This option is an up–and–out call, which expires worthless if the price of the underlying asset hits the barrier B from below, or has the same payoff at maturity as a call option with strike K otherwise.

The position of the strike K relative to the barrier B is also important. If the spot price $S(0)$ of the underlying asset at time 0 is such that the barrier is already triggered, then the option is either worth 0, or it is equivalent to a plain vanilla option, which is priced using the Black–Scholes formula. Therefore, we are interested in the case when the barrier is not trigerred already at time 0. There are sixteen different barrier options to price, corresponding to

$$\begin{array}{cccc} \textit{up} & \textit{in} & \textit{put} & B < K \\ \textit{down} & \textit{out} & \textit{call} & B \geq K \end{array}$$

Several barrier options can be priced by using simple no–arbitrage arguments, either in an absolute way, or relative to other barrier options.

For example, long positions in one up–and–out and one up–and–in option with all other parameters, i.e., barrier, maturity, and strike of the underlying option, being the same is equivalent to a long position in the corresponding plain vanilla option. It is easy to see that for every possible path for the price of the underlying asset, the final payoffs of the plain vanilla option and of a portfolio made of the up–and–out and the up–and–in options are the same:

If the barrier is hit before the expiry of the options, then the up–and–out option expires worthless, while the up–and–in option becomes a plain vanilla option. Thus, the payoff of the portfolio at maturity will be the same as the payoff of a plain vanilla option.

If the price of the underlying asset never reaches the barrier, then the up–and–in option is never knocked in, and will expire worthless, while the up–and–out option is never knocked out, and it will have the same payoff at maturity as a plain vanilla option. The payoff at maturity of the portfolio made of the barrier options will be the same as the payoff of a plain vanilla option.

Similarly, we can show that a portfolio made of a down–and–out and a down–and–in option is equivalent to a plain vanilla option.

This means that we only have to price either the "in" or the "out" option; the complementary option can be priced using the Black–Scholes formula for pricing plain vanilla options.

The following knock–out barrier options have value 0:
- up–and–out call with $S(0) < B \leq K$;
- down–and–out put with $S(0) > B > K$,

since the barrier must be triggered, and therefore the option will be knocked out, in order for the underlying option to expire in the money.

Using the "in"–"out" duality, we conclude that the following knock–in options have the same value as the corresponding plain vanilla options:

- the up–and–in call with $S(0) \leq B \leq K$;
- the down–and–in put with $S(0) > B > K$.

For all the options listed above, the barrier must be triggered in order for the option to expire in the money.

Therefore, it is enough to price the following barrier options:

1. up–and–in call with $B > K$;
2. down–and–in call with $B > K$;
3. down–and–in call with $B < K$;
4. up–and–in put with $B > K$;
5. up–and–in put with $B < K$;
6. down–and–in put with $B < K$.

Closed formulas for pricing these types of barrier options can be found in Haug [12]. For example, the value of the down–and–out call with $B < K$ is

$$V(S, K, t) \;=\; C(S, K, t) \;-\; \left(\frac{B}{S}\right)^{2a} C\left(\frac{B^2}{S}, K, t\right), \qquad (7.41)$$

where the constant a is given by formula (7.35), i.e., $a = \frac{r-q}{\sigma^2} - \frac{1}{2}$. Here, $C(S, K, t)$ is the value at time t of a plain vanilla call option with strike K on the same underlying asset, $C\left(\frac{B^2}{S}, K, t\right)$ is the Black–Scholes value at time t of a plain vanilla call with strike K on an asset having spot price $\frac{B^2}{S}$ (and the same volatility as the underlying). Note that, as expected, $V(S, K, t) < C(S, K, t)$. Also, $V(S, K, t) \approx C(S, K, t)$ as the barrier B goes to 0, when the probability of hitting the barrier and knocking the option out also goes to 0:

$$\lim_{B \searrow 0} V(S, t) \;=\; C(S, t).$$

As expected, closed formulas do not exist for pricing American barrier options. Numerical methods, e.g., binomial tree methods and finite differences, are used to price American barrier options. From the point of view of computational costs, the solutions of the numerical methods for barrier options are comparable to those for plain vanilla options, for finite difference methods, and more expensive for tree methods.

7.9 On the optimality of early exercise for American options

It is interesting to note that American call options on non–dividend paying assets are never optimal to exercise. Instead of exercising the call, entering into a static hedge by shorting one unit of the underlying asset is guaranteed to be more profitable. American calls on non–dividend paying assets are therefore worth the same as the corresponding European calls. This is not true for American put options, nor for American calls on assets paying dividends.

To see this, assume that, at time $t < T$, the call option is in the money, i.e., $S(t) > K$; here T denotes the maturity of the option. For the long call position (i.e., for the holder of the option), consider two different strategies:

First Strategy: Exercise the call at time t and deposit the premium $S(t) - K$ at the constant risk free rate r until time T. The value of the resulting portfolio at time T is

$$V_1(T) \;=\; (S(t) - K)\; e^{r(T-t)}; \tag{7.42}$$

Second Strategy: Do not exercise the call; short one unit of the underlying asset at time t, and wait until maturity T without exercising the American option. At time t, the portfolio will consist of:

- long 1 American call;
- short 1 unit of the underlying asset;
- cash $S(t)$.

The value of this portfolio at time T depends on whether the call expires in the money or not:

If $S(T) > K$, then the call is exercised at time T, and K is paid for one unit of the underlying asset, which is used to close the short position. The cash position of $S(t)$ at time t earns interest at the constant risk free rate r. Its value at time T is $S(t)e^{r(T-t)}$. The value of the portfolio will be

$$V_{2,1}(T) \;=\; S(t)\; e^{r(T-t)} - K. \tag{7.43}$$

If $S(T) \leq K$, then the call expires worthless. One unit is of the underlying asset is bought at time T, for the price $S(T)$, to close the short position. The value of the portfolio will be

$$V_{2,2}(T) \;=\; S(t)\; e^{r(T-t)} - S(T) \;\geq\; S(t)\; e^{r(T-t)} - K, \tag{7.44}$$

since $S(T) \leq K$ in this case.

From (7.43) and (7.44), we conclude that, if the second strategy is used, the value of the portfolio at time T is guaranteed to be at least

$$V_2(T) \;=\; \min(V_{2,1}(T), V_{2,2}(T)) \;=\; S(t)\; e^{r(T-t)} - K. \tag{7.45}$$

Note that

$$S(t)\; e^{r(T-t)} - K \;>\; (S(t) - K)\; e^{r(T-t)}, \tag{7.46}$$

provided that the interest rates are positive, i.e., that $r > 0$. From (7.42), (7.45), and (7.46) we conclude that the value of the second portfolio, i.e., of the static hedging strategy, is higher than the value of the portfolio if the option is exercised at time t before maturity, i.e., $V_2(T) > V_1(T)$.

In other words, it is never optimal to exercise early an American call option on a non–dividend paying asset. Shorting the underlying asset and holding the American call option until maturity will earn a greater payoff at maturity than exercising the call option early.

It is natural to ask why does the previous argument not hold if the underlying asset pays dividends, and for put options?

When shorting a dividend paying asset, you are responsible for paying the dividends to the lender of the asset. Therefore, if the static hedge strategy is employed, the value of the portfolio at time T will be reduced by the amount of the dividends paid. Then, if the call option expires in the money, all we can say[7] is that

$$V_{2,1}(T) \;<\; S(t)\; e^{r(T-t)} - K,$$

which no longer means that $V_1(T) = (S(t) - K)e^{r(T-t)} < V_{2,1}(T)$. In other words, the static hedging strategy of shorting the underlying asset is no longer guaranteed to be worth more at time T than the strategy of exercising the call at time t.

For an American put option which is in the money at time t, i.e., $S(t) < K$, exercising the option and holding the cash until maturity T generates the following payoff at time T:

$$V_3(T) \;=\; (K - S(t))\; e^{r(T-t)}.$$

The static hedging strategy for the long put position is to buy one unit of the underlying asset (paying $S(t)$ cash to do so). At time t, the portfolio will consist of long 1 American put, long 1 unit of the underlying asset, and $-S(t)$ in cash. In this case, the interest earned by the cash borrowed to buy the asset reduces the value of the portfolio. (This is similar to the role played by dividends for the static hedging of American call options.)

If the put expires in the money, i.e., if $S(T) < K$, then it is exercised, and the share will be sold for K. The value of the portfolio in this case is

$$V_{4,1}(T) \;=\; K - S(t)\; e^{r(T-t)}.$$

[7] If the underlying asset pays dividends continuously at the rate q, it can be shown that $V_{2,1}(T) = S(t)\; e^{(r-q)(T-t)} - K$.

If the put expire worthless, i.e., if $S(T) \geq K$, the value of the portfolio is

$$V_{4,2}(T) = S(T) - S(t)\ e^{r(T-t)} \geq K - S(t)\ e^{r(T-t)} = V_{4,1}(T).$$

Thus,

$$V_4(T) = \min(V_{4,1}(T), V_{4,2}(T)) = K - S(t)\ e^{r(T-t)}.$$

For early exercise not to be optimal for the American put, the value $V_4(T)$ of the static hedging portfolio at time T should be greater than $V_3(T)$, for any value $S(T)$ of the underlying asset at maturity. However,

$$V_4(T) = K - S(t)\ e^{r(T-t)} < (K - S(t))\ e^{r(T-t)} = V_3(T),$$

and the argument that worked for the American call option does not work for put options.

We note that an American put option is optimal to exercise if it is deep enough in the money. For example,in section 1.8, we showed that the value at time t of a European put option on an underlying asset with spot price equal to 0, i.e., with $S(t) = 0$, is $P_{Eur}(t) = Ke^{-r(T-t)}$; cf. (1.45). However, the value of the corresponding American option is $P_{Amer}(t) = K$, since the option is exercised when the price of the underlying is 0 for the intrinsic payoff $K - S(t) = K$. Therefore, $P_{Amer}(t) > P_{Eur}(t)$, and early exercise is optimal in this case. Finding the early exercise region of an American put option is a more subtle question that we will not address here.

7.10 References

Different presentations of the multivariable calculus topics discussed in this section can be found in [10] and [20].

A complete list of closed formulas for barrier options can be found in Haug [12]. The monograph by Glasserman [11] discusses Monte Carlo methods with applications in financial engineering. The derivation and solution of the Black–Scholes PDE, as well as its extensions to futures (Black's formula) and time–dependent parameters are presented in Wilmott et al. [35].

7.11 Exercises

1. For $t = 0$ and $q = 0$, the formula (3.68) for the Gamma of a plain vanilla European call option reduces to

$$\Gamma(C) = \frac{1}{S\sigma\sqrt{2\pi T}} \exp\left(-\frac{d_1^2}{2}\right),$$

where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

Show that, as a function of $S > 0$, the Gamma of the call option is first increasing until it reaches a maximum point and then decreases. Also, show that

$$\lim_{S\to 0} \Gamma(S) = 0 \quad \text{and} \quad \lim_{S\to\infty} \Gamma(S) = 0.$$

2. Let D be the domain bounded by the x–axis, the y–axis, and the line $x + y = 1$. Compute

$$\int\int_D \frac{x-y}{x+y}\, dxdy.$$

Hint: Use the change of variables $s = x + y$ and $t = x - y$, which is equivalent to $x = \frac{s+t}{2}$ and $y = \frac{s-t}{2}$. Note that $(x, y) \in D$ if and only if $0 \leq s \leq 1$ and $-s \leq t \leq s$.

3. Use the change of variables to polar coordinates to show that the area of a circle of radius R is πR^2, i.e., prove that

$$\int\int_{D(0,R)} 1\ dxdy = \pi R^2.$$

4. Let $V(S,t) = \exp(-ax - b\tau)u(x,\tau)$, where

$$x = \ln\left(\frac{S}{K}\right), \quad \tau = \frac{(T-t)\sigma^2}{2}, \quad a = \frac{r-q}{\sigma^2} - \frac{1}{2}, \quad b = \left(\frac{r-q}{\sigma^2} + \frac{1}{2}\right)^2.$$

This is the change of variables that reduces the Black–Scholes PDE for $V(S,t)$ to the heat equation for $u(x,\tau)$.

(i) Show that the boundary condition $V(S,T) = \max(S - K, 0)$ for the European call option becomes the following boundary condition for $u(x,\tau)$ at time $\tau = 0$:

$$u(x,0) = K\ \exp(ax)\ \max(e^x - 1, 0).$$

(ii) Show that the boundary condition $V(S,T) = \max(K - S, 0)$ for the European put option becomes

$$u(x,0) = K \exp(ax) \max(1 - e^x, 0).$$

5. Show that the solution of the system

$$\begin{cases} 2a + 1 - \frac{2(r-q)}{\sigma^2} = 0 \\ b + a^2 + a\left(1 - \frac{2(r-q)}{\sigma^2}\right) - \frac{2r}{\sigma^2} = 0 \end{cases}$$

is $a = \frac{r-q}{\sigma^2} - \frac{1}{2}$ and $b = \left(\frac{r-q}{\sigma^2} + \frac{1}{2}\right)^2$.
Note: An elegant way to do this is to notice that

$$\begin{aligned} & b + a^2 - 2a\left(\frac{r-q}{\sigma^2} - \frac{1}{2}\right) - \frac{2(r-q)}{\sigma^2} \\ = \; & b + \left(a - \left(\frac{r-q}{\sigma^2} - \frac{1}{2}\right)\right)^2 - \left(\frac{r-q}{\sigma^2} - \frac{1}{2}\right)^2 - \frac{2(r-q)}{\sigma^2} \\ = \; & b + \left(a - \left(\frac{r-q}{\sigma^2} - \frac{1}{2}\right)\right)^2 - \left(\frac{r-q}{\sigma^2} + \frac{1}{2}\right)^2. \end{aligned}$$

6. Assume that the function $V(S,I,t)$ satisfies the following PDE:

$$\frac{\partial V}{\partial t} + S\frac{\partial V}{\partial I} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0. \tag{7.47}$$

Consider the following change of variables:

$$V(S,I,t) = S\, H(R,t), \quad \text{where} \quad R = \frac{I}{S}. \tag{7.48}$$

Show that $H(R,t)$ satisfies the following PDE:

$$\frac{\partial H}{\partial t} + \frac{1}{2}\sigma^2 R^2 \frac{\partial^2 H}{\partial R^2} + (1 - rR)\frac{\partial H}{\partial R} = 0. \tag{7.49}$$

Note: An Asian call option pays the maximum between the spot price $S(T)$ of the underlying asset at maturity T and the average price of the underlying asset over the entire life of the option, i.e.,

$$\max\left(S(T) - \frac{1}{T}\int_0^T S(\tau)\, d\tau\right).$$

Thus, the price $V(S,I,t)$ of an Asian option depends not only on the spot price S of the underlying asset and on the time t, but also on the average price $I(t)$ of the underlying between time 0 and time t, where

$$I(t) = \int_0^t S(\tau)\, d\tau.$$

It can be shown that $V(S,I,t)$ satisfies the PDE (7.47). Similarity solutions of the type (7.48) are good candidates for solving the PDE (7.47). The PDE (7.49) satisfied by $H(R,T)$ can be solved numerically, e.g., by using finite differences.

7. The price of a non-dividend-paying asset is lognormally distributed. Assume that the spot price is 40, the volatility is 30%, and the interest rates are constant at 5%. Fill in the Black–Scholes values of the OTM put options on the asset in the table below:

Option Type	Strike	Maturity	Value
Put	45	6 months	
Put	45	3 months	
Put	48	6 months	
Put	48	3 months	
Put	51	6 months	
Put	51	3 months	

For which of these options is the intrinsic value $\max(K - S, 0)$ larger than the price of the option (in which case the corresponding American put is guaranteed to be worth more than the European put)?

8. Show that the premium of the price of a European call option over its intrinsic value $\max(S - K, 0)$ is largest at the money. In other words, show that the maximum value of

$$C_{BS}(S) - \max(S - K, 0)$$

is obtained for $S = K$, where $C_{BS}(S)$ is the Black–Scholes value of the plain vanilla European call option with strike K and spot price S.

9. Use formula (7.41) to price a six months down–and–out call on a non–dividend–paying asset with price following a lognormal distribution with 30% volatility and spot price 40. The barrier is $B = 35$ and the strike for the call is $K = 40$. The risk–free interest rate is constant at 5%.

10. Show that the value of a down–and–out call with barrier B less than the strike K of the call, i.e., $B < K$, converges to the value of a plain vanilla call with strike K when $B \searrow 0$. For simplicity, assume that the underlying asset does not pay dividends and that interest rates are zero.

    Hint: Use formula (7.41) to show that the value of the down–and–out call is

$$V(S,t) = C(S,t) - \left(BN(d_1) - \frac{SK}{B} e^{-rT} N(d_2) \right),$$

    where $C(S,t)$ is the value of the plain vanilla call with strike K and

$$d_1 = \frac{\ln\left(\frac{B^2}{SK}\right) + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = \frac{\ln\left(\frac{B^2}{SK}\right) - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}.$$

    Use l'Hôpital's rule to show that

$$\lim_{B\searrow 0} \frac{1}{B} N(d_2) = 0.$$

11. Compute the Delta and Gamma of a down–and–out call with $B < K$, using formula (7.41).

Chapter 8

Lagrange multipliers. $N-$ dimensional Newton's method. Implied volatility. Bootstrapping.

The Lagrange multipliers method for finding absolute extrema of multivariable functions.

Newton's method, bisection method, and secant method for solving one dimensional nonlinear problems.

Newton's method for solving N–dimensional nonlinear problems.

8.1 Lagrange multipliers

Optimization problems often require finding extrema of multivariable functions subject to various constraints. One method to solve such problems is by using Lagrange multipliers, as outlined below.

Let $U \subset \mathbb{R}^n$ be an open set, and let $f : U \to \mathbb{R}$ be a smooth function, e.g., infinitely many times differentiable. We want to find the extrema of $f(x)$ subject to m constrains given by $g(x) = 0$, where $g : U \to \mathbb{R}^m$ is a smooth function, i.e.,

Find $x_0 \in U$ such that

$$\max_{\substack{g(x)=0 \\ x \in U}} f(x) = f(x_0) \quad \text{or} \quad \min_{\substack{g(x)=0 \\ x \in U}} f(x) = f(x_0). \tag{8.1}$$

Problem (8.1) is called a constrained optimization problem. For this problem to be well posed, a natural assumption is that the number of constraints is smaller than the number of degrees of freedom, i.e., $m < n$.

Another way to formalize problem (8.1) is to introduce the set S of points satisfying the constraint $g(x) = 0$, and find extrema for the restriction of the function $f(x)$ to the space S, i.e., for $f|_S$.

Definition 8.1. *The point $x_0 \in U \subset \mathbb{R}^n$ is called a constrained extremum of the function $f : U \to \mathbb{R}$ with respect to the constraint $g(x) = 0$ if and only if x_0 is an extremum point of $f|_S$, where $S = \{x \in U$ such that $g(x) = 0\}$.*

To solve the constrained optimization problem (8.1), let $\lambda = (\lambda_i)_{i=1:m}$ be a vector of the same size, m, as the number of constraints; λ is called a Lagrange multipliers vector. Let $F : U \times \mathbb{R}^m \to \mathbb{R}$ given by

$$F(x, \lambda) \;=\; f(x) \;+\; \lambda^t\, g(x) \tag{8.2}$$

be the Lagrangian function of f and g. If

$$g(x) \;=\; \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{pmatrix},$$

then $F(x, \lambda)$ can be written explicitly as

$$F(x, \lambda) \;=\; f(x) \;+\; \sum_{i=1}^{m} \lambda_i g_i(x).$$

It is easy to see that

$$\frac{\partial F}{\partial x_j}(x, \lambda) \;=\; \frac{\partial f}{\partial x_j}(x) \;+\; \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j}(x), \quad \forall\, j = 1 : n; \tag{8.3}$$

$$\frac{\partial F}{\partial \lambda_i}(x, \lambda) \;=\; g_i(x), \quad \forall\, i = 1 : m. \tag{8.4}$$

The gradient[1] of $F(x, \lambda)$ with respect to both x and λ will be denoted by $\nabla_{(x,\lambda)} F(x, \lambda)$, and is the following row vector:

$$\nabla_{(x,\lambda)}\, F(x, \lambda) \;=\; \left(\; \nabla_x\, F(x, \lambda) \;\; \nabla_\lambda\, F(x, \lambda) \;\right); \tag{8.5}$$

cf. (1.36). From (8.3) and (8.4) it follows that

$$\nabla_x\, F(x, \lambda) \;=\; \left(\frac{\partial F}{\partial x_1}(x, \lambda) \;\ldots\; \frac{\partial F}{\partial x_n}(x, \lambda) \right) = \nabla f(x) + (\nabla g(x))^t \lambda; \tag{8.6}$$

$$\nabla_\lambda\, F(x, \lambda) \;=\; \left(\frac{\partial F}{\partial \lambda_1}(x, \lambda) \;\ldots\; \frac{\partial F}{\partial \lambda_m}(x, \lambda) \right) = g(x), \tag{8.7}$$

[1] In this section, we use the notation ∇F, instead of DF, for the gradient of F.

where $\nabla f(x)$ and $\nabla g(x)$, the gradients of $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}^m$ are given by

$$\nabla f(x) \;=\; \left(\frac{\partial f}{\partial x_1}(x, \lambda) \;\ldots\; \frac{\partial f}{\partial x_n}(x, \lambda) \right);$$

$$\nabla g(x) \;=\; \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \frac{\partial g_1}{\partial x_2}(x) & \ldots & \frac{\partial g_1}{\partial x_n}(x) \\ \frac{\partial g_2}{\partial x_1}(x) & \frac{\partial g_2}{\partial x_2}(x) & \ldots & \frac{\partial g_2}{\partial x_n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x) & \frac{\partial g_m}{\partial x_2}(x) & \ldots & \frac{\partial g_m}{\partial x_n}(x) \end{pmatrix};$$

cf. (1.36) and (1.38), respectively. From (8.5–8.7), we conclude that

$$\nabla_{(x,\lambda)}\, F(x, \lambda) \;=\; \left(\; \nabla f(x) + (\nabla g(x))^t \lambda \quad g(x) \;\right). \tag{8.8}$$

In the Lagrange multipliers method, the constrained extremum point x_0 is found by identifying the critical points of the Lagrangian $F(x, \lambda)$. For the method to work, the following condition must be satisfied:

The gradient $\nabla g(x)$ has full rank at any point x in the constrained space S, i.e.,

$$\mathit{rank}(\nabla g(x)) \;=\; m, \quad \forall\; x \in S. \tag{8.9}$$

The following theorem gives necessary conditions for a point $x_0 \in U$ to be a constrained extremum point for $f(x)$. Its proof involves the inverse function theorem and is beyond the scope of this book.

Theorem 8.1. *Assume that the constraint function $g(x)$ satisfies the condition (8.9). If $x_0 \in U$ is a constrained extremum of $f(x)$ with respect to the constraint $g(x) = 0$, then there exists a Lagrange multiplier $\lambda_0 \in \mathbb{R}^m$ such that the point (x_0, λ_0) is a critical point for the Lagrangian function $F(x, \lambda)$, i.e., such that*

$$\nabla_{(x,\lambda)}\, F(x_0, \lambda_0) \;=\; 0. \tag{8.10}$$

We note that $\nabla_{(x,\lambda)}\, F(x, \lambda)$ is a function from $\mathbb{R}^{m+n}$ into $\mathbb{R}^{m+n}$. Thus, solving (8.10) to find the critical points of $F(x, \lambda)$ requires, in general, using the N–dimensional Newton's method for solving nonlinear equations. Interestingly enough, for some practical problems such as finding efficient portfolios, problem (8.10) is actually a linear system which can be solved without using Newton's method; see section 8.4 for details.

If condition (8.9) is not satisfied, then Theorem 8.1 does not necessarily hold, as seen in the example below:

Example: Find the minimum of $x_1^2 + x_1 + x_2^2$, subject to the constraint $(x_1 - x_2)^2 = 0$.

Answer: This problem can be solved without using Lagrange multipliers as follows: If $(x_1 - x_2)^2 = 0$, then $x_1 = x_2$. The function to minimize becomes

$$x_1^2 + x_1 + x_2^2 \;=\; 2x_1^2 + x_1 \;=\; 2\left(x_1 + \frac{1}{4}\right)^2 - \frac{1}{8},$$

which achieves its minimum when $x_1 = -\frac{1}{4}$. We conclude that there exists a unique constrained minimum point $x_1 = x_2 = -\frac{1}{4}$, and that the corresponding minimum value is $-\frac{1}{8}$.

Although we solved the problem directly, we attempt to find an alternative solution using the Lagrange multipliers method. Using the framework outlined above, we want to find the minimum of the function

$$f(x_1, x_2) \;=\; x_1^2 + x_1 + x_2^2$$

for $(x_1, x_2) \in \mathbb{R}^2$, such that the constraint $g(x_1, x_2) = 0$ is satisfied, where

$$g(x_1, x_2) \;=\; (x_1 - x_2)^2.$$

Note that $g(x_1, x_2) = 0$ if and only if $x_1 = x_2$, and let $S = \{(x_1, x_2) \in \mathbb{R}^2$ such that $x_1 = x_2\}$ be the space where the constraint $g(x_1, x_2) = 0$ is satisfied. By a simple computation,

$$\nabla g(x) \;=\; \begin{pmatrix} \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} \end{pmatrix} \;=\; \begin{pmatrix} 2(x_1 - x_2) & -2(x_1 - x_2) \end{pmatrix}.$$

For any $x = (x_1, x_2) \in S$, it follows that $x_1 = x_2$. Therefore, $\nabla g(x) = (0\ 0)$ and $\operatorname{rank}(\nabla g(x)) = 0$, for all $x \in S$. We conclude that condition (8.9) is not satisfied at any point $x = (x_1, x_2) \in S$.

Since the problem has one constraint, we only have one Lagrange multiplier, denoted by $\lambda \in \mathbb{R}$. From (8.2), we find that the Lagrangian is

$$F(x_1, x_2, \lambda) \;=\; f(x_1, x_2) \;+\; \lambda g(x_1, x_2) \;=\; x_1^2 + x_1 + x_2^2 \;+\; \lambda(x_1 - x_2)^2.$$

The gradient $\nabla_{(x,\lambda)} F(x, \lambda)$ is computed as in (8.8) and has the form

$$\nabla_{(x,\lambda)}\, F(x, \lambda) = \begin{pmatrix} 2x_1 + 1 + 2\lambda(x_1 - x_2) & 2x_2 - 2\lambda(x_1 - x_2) & (x_1 - x_2)^2 \end{pmatrix}.$$

Finding the critical points for $F(x, \lambda)$ requires solving $\nabla_{(x,\lambda)}\, F(x, \lambda) = 0$, which is equivalent to the following system of equations:

$$\begin{cases} 2x_1 + 1 + 2\lambda(x_1 - x_2) &= 0; \\ 2x_2 - 2\lambda(x_1 - x_2) &= 0; \\ (x_1 - x_2)^2 &= 0. \end{cases}$$

This system does not have a solution: from the third equation, we obtain that $x_1 = x_2$. Then, from the second equation, we find that $x_2 = 0$, which implies that $x_1 = 0$. Substituting $x_1 = x_2 = 0$ into the first equation, we obtain $1 = 0$, which is a contradiction.

In other words, the Lagrangian $F(x, \lambda)$ has no critical points. However, we showed before that the point $(x_1, x_2) = \left(-\frac{1}{4}, -\frac{1}{4}\right)$ is a constrained minimum point for $f(x)$ given the constraint $g(x) = 0$. The reason Theorem 8.1 does not apply in this case is that assumption (8.9), which was required for Theorem 8.1 to hold, is not satisfied. □

Finding sufficient conditions for a critical point (x_0, λ_0) for the Lagrangian function $F(x, \lambda)$ to correspond to a constrained extremum point x_0 for $f(x)$ is somewhat more complicated (and rather rarely checked in practice, although that should not be the case).

Consider the function $F_0 : U \to \mathbb{R}$ given by

$$F_0(x) \;=\; F(x, \lambda_0) \;=\; f(x) \;+\; \lambda_0^t g(x).$$

Let $D^2 F_0(x_0)$ be the Hessian of $F_0(x)$ evaluated at the point x_0, i.e.,

$$D^2 F_0(x_0) \;=\; \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2}(x_0, \lambda_0) & \frac{\partial^2 F}{\partial x_2 \partial x_1}(x_0, \lambda_0) & \cdots & \frac{\partial^2 F}{\partial x_n \partial x_1}(x_0, \lambda_0) \\ \frac{\partial^2 F}{\partial x_1 \partial x_2}(x_0, \lambda_0) & \frac{\partial^2 F}{\partial x_2^2}(x_0, \lambda_0) & \cdots & \frac{\partial^2 F}{\partial x_n \partial x_2}(x_0, \lambda_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_1 \partial x_n}(x_0, \lambda_0) & \frac{\partial^2 F}{\partial x_2 \partial x_n}(x_0, \lambda_0) & \cdots & \frac{\partial^2 F}{\partial x_n^2}(x_0, \lambda_0); \end{pmatrix} \tag{8.11}$$

cf. (1.37). Note that $D^2 F_0(x_0)$ is an $n \times n$ matrix.

Let $q(v)$ be the quadratic form associated to the matrix $D^2 F_0(x_0)$, i.e.,

$$q(v) \;=\; v^t\ D^2 F_0(x_0)\ v \;=\; \sum_{1 \le i,j \le n} \frac{\partial^2 F}{\partial x_i\, \partial x_j}(x_0, \lambda_0) v_i v_j, \tag{8.12}$$

with $v = (v_i)_{i=1:n}$.

We restrict our attention to the vectors v satisfying $\nabla g(x_0)\ v = 0$, i.e., to the vector space

$$V_0 \;=\; \{v \in \mathbb{R}^n \mid \nabla g(x_0)\ v = 0\}. \tag{8.13}$$

If condition (8.9) is satisfied, it follows that $\operatorname{rank}(\nabla g(x_0)) = m$. Assume, without losing any generality, that the first m columns of $\nabla g(x_0)$ are linearly independent, and let

$$v_{red} \;=\; \begin{pmatrix} v_{m+1} \\ \vdots \\ v_n \end{pmatrix}.$$

If we formally solve the linear system $\nabla g(x_0)\ v = 0$, we obtain that the entries $v_1, v_2, \ldots, v_m$ of the vector v can be written as linear combinations

of $v_{m+1}, v_{m+2}, \ldots, v_n$, the other $n-m$ entries of v. Then, by restricting $q(v)$ to the vector space V_0, we can write $q(v)$ as a quadratic form depending only on the entries of the vector v_{red}, i.e.,

$$q_{red}(v_{red}) = q(v) = \sum_{m+1 \leq i,j \leq n} q_{red}(i,j) v_i v_j, \quad \forall\, v \in V_0. \tag{8.14}$$

Example: For clarification purposes, we provide an example of deriving the reduced quadratic form $q_{red}(v_{red})$ from $q(v)$ for $n=3$ and $m=1$.

Assume that

$$D_x^2 F(x_0, \lambda_0) = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 2 \\ 0 & 2 & 3 \end{pmatrix} \quad \text{and} \quad \nabla g(x_0) = (\, 1 \;\; -1 \;\; 2 \,).$$

If $v = (v_i)_{i=1:3}$, then

$$q(v) = v^t\, D_x^2 F(x_0, \lambda_0)\, v = v_1^2 + 2v_2^2 + 3v_3^2 - 2v_1 v_2 + 4v_2 v_3.$$

The condition $\nabla g(x_0) v = 0$ is equivalent to $v_1 - v_2 + 2v_3 = 0$. Then, $v_1 = v_2 - 2v_3$ and

$$\begin{aligned} q(v) &= (v_2 - 2v_3)^2 + 2v_2^2 + 3v_3^2 - 2(v_2 - 2v_3)v_2 + 4v_2 v_3 \\ &= v_2^2 + 4v_2 v_3 + 7v_3^2. \end{aligned}$$

Note that $v_{red} = \begin{pmatrix} v_2 \\ v_3 \end{pmatrix}$. The reduced quadratic form $q_{red}(v_{red})$ is

$$q_{red}(v_{red}) = q(v) = v_2^2 + 4v_2 v_3 + 7v_3^2. \quad \Box \tag{8.15}$$

Whether the point x_0 is a constrained extremum for $f(x)$ will depend on the quadratic form q_{red} being either positive definite[2], i.e.,

$$q_{red}(v_{red}) > 0, \quad \forall\, v_{red} \neq 0,$$

or negative definite, i.e.,

$$q_{red}(v_{red}) < 0, \quad \forall\, v_{red} \neq 0.$$

This result is presented in the following theorem:

[2]For example, the quadratic form q_{red} from (8.15) is positive definite since

$$q_{red}(v_{red}) = (v_2 + 2v_3)^2 + 3v_3^2, \quad \forall\, (v_2, v_3) \neq (0,0).$$

Theorem 8.2. *Assume that the constraint function $g(x)$ satisfies condition (8.9). Let $x_0 \in U \subset \mathbb{R}^n$ and $\lambda_0 \in \mathbb{R}^m$ such that the point (x_0, λ_0) is a critical point for the Lagrangian function $F(x, \lambda) = f(x) + \lambda_0^t g(x)$.*

If the reduced quadratic form (8.14) corresponding to the point (x_0, λ_0) is positive definite, then x_0 is a constrained minimum for $f(x)$ with respect to the constraint $g(x) = 0$.

If the reduced quadratic form (8.14) is negative definite, then x_0 is a constrained maximum for $f(x)$ with respect to the constraint $g(x) = 0$.

We note that the conditions from Theorem 8.2 are, in practice, rarely checked. While this is not recommended, for some financial applications it may be sound to do so. The reason is that if a problem is known to have a unique constrained extremum, and if the Lagrangian has exactly one critical point, then that critical point must be the constrained extremum point. For this line of thinking to be sound we must know that the constrained extremum problem has a unique solution. In general, showing that a problem has a unique solution is not straightforward.

The steps required to solve a constrained optimization problem using Lagrange multipliers can be summarized as follows:

Step 1: Check that $\text{rank}(\nabla g(x)) = m$, for all $x \in S$.

Step 2: Find $(x_0, \lambda_0) \in U \times \mathbb{R}^m$ such that $\nabla_{(x,\lambda)} F(x_0, \lambda_0) = 0$.

Step 3.1: Compute $q(v) = v^t\, D^2 F_0(x_0)\, v$, where $F_0(x) = f(x) + \lambda_0^t g(x)$.

Step 3.2: Compute $q_{red}(v_{red})$ by restricting $q(v)$ to vectors v satisfying the condition $\text{rank}(\nabla g(x_0))v = 0$. Decide whether $q_{red}(v_{red})$ is positive definite or negative definite.

Step 4: Use Theorem 8.2 to decide whether x_0 is a constrained minimum point or a constrained maximum point.

We include two examples below. An application of Lagrange multipliers to portfolio optimization is presented in section 8.4.

Example: Find the positive numbers x_1, x_2, x_3 such that $x_1 x_2 x_3 = 1$ and $x_1 x_2 + x_2 x_3 + x_3 x_1$ is minimized.

Answer: We first reformulate the problem as a constrained optimization problem. Let $U = \prod_{i=1:3}(0, \infty) \subset \mathbb{R}^3$ and let $x = (x_1, x_2, x_3) \in U$. The functions $f: U \to \mathbb{R}$ and $g: U \to \mathbb{R}$ are defined as

$$f(x) = x_1 x_2 + x_2 x_3 + x_3 x_1; \quad g(x) = x_1 x_2 x_3 - 1.$$

We want to minimize $f(x)$ over the set U subject to the constraint $g(x) = 0$.

Step 1: Check that $\text{rank}(\nabla g(x)) = 1$ for any x such that $g(x) = 0$.

Let $x = (x_1, x_2, x_3) \in U$. It is easy to see that

$$\nabla g(x) = (x_2x_3 \;\; x_1x_3 \;\; x_1x_2). \tag{8.16}$$

Note that $\nabla g(x) \neq 0$, since $x_i > 0$, $i = 1:3$. Therefore, $\text{rank}(\nabla g(x)) = 1$ for all $x \in U$, and condition (8.9) is satisfied.

Step 2: Find (x_0, λ_0) such that $\nabla_{(x,\lambda)} F(x_0, \lambda_0) = 0$.
The Lagrangian associated to this problem is

$$F(x, \lambda) = x_1x_2 + x_2x_3 + x_3x_1 + \lambda\,(x_1x_2x_3 - 1), \tag{8.17}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier.

Let $x_0 = (x_{0,1}, x_{0,2}, x_{0,3})$. From (8.17), we find that $\nabla_{(x,\lambda)} F(x_0, \lambda_0) = 0$ is equivalent to the following system:

$$\begin{cases} x_{0,2} + x_{0,3} + \lambda_0\, x_{0,2}\, x_{0,3} = 0; \\ x_{0,1} + x_{0,3} + \lambda_0\, x_{0,1}\, x_{0,3} = 0; \\ x_{0,1} + x_{0,2} + \lambda_0\, x_{0,1}\, x_{0,2} = 0; \\ x_{0,1}\, x_{0,2}\, x_{0,3} = 1. \end{cases}$$

By multiplying the first three equations by $x_{0,1}$, $x_{0,2}$, and $x_{0,3}$, respectively, and using the fact that $x_{0,1}x_{0,2}x_{0,3} = 1$, we obtain that

$$-\lambda = x_{0,1}\, x_{0,2} + x_{0,1}\, x_{0,3} = x_{0,1}\, x_{0,2} + x_{0,2}\, x_{0,3} = x_{0,1}\, x_{0,3} + x_{0,2}\, x_{0,3}.$$

Since $x_{0,i} \neq 0$, $i = 1:3$, we find that $x_{0,1} = x_{0,2} = x_{0,3}$. Since $x_{0,1}\, x_{0,2}\, x_{0,3} = 1$, we conclude that $x_{0,1} = x_{0,2} = x_{0,3} = 1$ and $\lambda_0 = -2$.

Step 3.1: Compute $q(v) = v^t\, D^2F_0(x_0)\, v$.
Since $\lambda_0 = -2$, we find that $F_0(x) = f(x) + \lambda_0^t g(x)$ is given by

$$F_0(x) = x_1x_2 + x_2x_3 + x_3x_1 - 2x_1x_2x_3 + 2.$$

The Hessian of $F_0(x)$ is

$$D^2F_0(x_1, x_2, x_3) = \begin{pmatrix} 0 & 1 - 2x_3 & 1 - 2x_2 \\ 1 - 2x_3 & 0 & 1 - 2x_1 \\ 1 - 2x_2 & 1 - 2x_1 & 0 \end{pmatrix},$$

and therefore

$$D^2F_0(1, 1, 1) = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}.$$

From (8.12), we find that the quadratic form $q(v)$ is

$$q(v) = v^t\, D^2F_0(1, 1, 1)\, v = -2v_1v_2 - 2v_2v_3 - 2v_1v_3. \tag{8.18}$$

Step 3.2: Compute $q_{red}(v_{red})$.
We first solve formally the equation $\nabla g(1,1,1)\; v = 0$, where $v = (v_i)_{i=1:3}$ is an arbitrary vector. From (8.16), we find that $\nabla g(1,1,1) = (1\; 1\; 1)$, and

$$\nabla g(1,1,1)\; v = v_1 + v_2 + v_3 = 0.$$

By solving for v_1 in terms of v_2 and v_3 we obtain that $v_1 = -v_2 - v_3$. Let $v_{red} = \begin{pmatrix} v_2 \\ v_3 \end{pmatrix}$. We substitute $-v_2 - v_3$ for v_1 in (8.18), to obtain the reduced quadratic form $q_{red}(v_{red})$, i.e.,

$$q_{red}(v_{red}) = 2v_2^2 + 2v_2v_3 + 2v_3^2 = v_2^2 + v_3^2 + (v_2 + v_3)^2.$$

Therefore, $q_{red}(v_{red}) > 0$ for all $(v_2, v_3) \neq (0, 0)$, which means that q_{red} is a positive definite form.

Step 4: From Theorem 8.2, we conclude that the point $x_0 = (1, 1, 1)$ is a constrained minimum for the function $f(x) = x_1x_2 + x_2x_3 + x_3x_1$, with $x_1, x_2, x_3 > 0$, subject to the constraint $x_1x_2x_3 = 1$. □

Example: Show that

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}, \quad \forall\, x_i > 0,\; i = 1:n. \tag{8.19}$$

This is the arithmetic–geometric[3] means inequality: the left hand side of (8.19) is the arithmetic mean of $(x_i)_{i=1:n}$, and the right hand side of (8.19) is the geometric mean of $(x_i)_{i=1:n}$.

Answer: We first note that inequality (8.19) is scalable in the sense that, if each variable x_i is multiplied by the same constant $c > 0$, the inequality (8.19) remains the same. We can therefore assume, without any loss of generality, that $\prod_{i=1}^{n} x_i = 1$, and inequality (8.19) reduces to proving that

$$\sum_{i=1}^{n} x_i \geq n, \quad \forall\, x_i > 0,\; i = 1:n, \quad \text{with} \quad \prod_{i=1}^{n} x_i = 1. \tag{8.20}$$

[3]It can be shown that the geometric–harmonic means inequality

$$\left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}} \geq \frac{n}{\sum_{i=1}^{n} \frac{1}{y_i}}, \quad \forall\, y_i > 0,\; i = 1:n$$

follows immediately from (8.19) by substituting $x_i = \frac{1}{y_i}$, $i = 1:n$.

We formulate problem (8.20) as a constrained optimization problem. Let $U = \prod_{i=1}^{n}(0,\ \infty)^n$ and let $x = (x_1, x_2, \ldots, x_n) \in U$. The functions $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ are defined as

$$f(x) \;=\; \sum_{i=1}^{n} x_i; \quad g(x) \;=\; \prod_{i=1}^{n} x_i \;-\; 1.$$

We want to minimize $f(x)$ over the set U subject to the constraint $g(x) = 0$.
Step 1: Check that $\mathrm{rank}(\nabla g(x)) = 1$ for any x such that $g(x) = 0$.
By direct computation, we find that

$$\nabla g(x) \;=\; \left(\prod_{i\neq 1} x_i \;\; \prod_{i\neq 2} x_i \;\; \ldots \;\; \prod_{i\neq n} x_i \right). \tag{8.21}$$

Note that $\nabla g(x) \neq 0$ since $x_i > 0$, $i = 1 : n$, for all $x \in U$. Therefore, $\mathrm{rank}(\nabla g(x)) = 1$ for any $x \in U$.
Step 2: Find (x_0, λ_0) such that $\nabla_{(x,\lambda)} F(x_0, \lambda_0) = 0$.
The Lagrangian associated to this problem is

$$F(x, \lambda) \;=\; \sum_{i=1}^{n} x_i \;+\; \lambda \left(\prod_{i=1}^{n} x_i - 1 \right), \tag{8.22}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. Let $x_0 = (x_{0,i})_{i=1:n} \in U$ and $\lambda_0 \in \mathbb{R}$. From (8.22), we find that $\nabla_{(x,\lambda)} F(x_0, \lambda_0) = 0$ can be written as

$$\begin{cases} 1 \;+\; \lambda_0 \prod_{i\neq j} x_{0,i} \;=\; 0, & \forall\, j = 1 : n; \\ \prod_{i=1}^{n} x_{0,i} \;=\; 1, & \end{cases}$$

which is the same as

$$\begin{cases} 1 \;+\; \frac{\lambda_0}{x_{0,j}} \;=\; 0, & \forall\, j = 1 : n; \\ \prod_{i=1}^{n} x_{0,i} \;=\; 1. & \end{cases}$$

The only solution to this system satisfying $x_{0,i} > 0$ for all $i = 1 : n$ is $x_0 = (1, 1, \ldots, 1)$ and $\lambda_0 = -1$.
Step 3.1: Compute $q(v) = v^t\, D^2 F_0(x_0)\, v$.
Since $\lambda_0 = -1$, we find that $F_0(x) = f(x) + \lambda_0^t g(x)$ is given by

$$F_0(x) \;=\; \sum_{i=1}^{n} x_i \;-\; \prod_{i=1}^{n} x_i + 1.$$

Then,

$$D^2 F_0(1, 1, \ldots, 1) \;=\; \begin{pmatrix} 0 & -1 & -1 & \ldots & -1 \\ -1 & 0 & -1 & \ldots & -1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -1 & \ldots & \ldots & -1 & 0 \end{pmatrix}.$$

The quadratic form $q(v) = v^t D^2 F_0(1, 1, \ldots, 1) v$ can be written as

$$\begin{aligned} q(v) \;&=\; -\sum_{1\leq i\neq j\leq n} v_i v_j \;=\; -2 \sum_{1\leq i<j\leq n} v_i v_j \\ &=\; -2v_1 \left(\sum_{i=2}^{n} v_i \right) \;-\; 2 \sum_{2\leq i<j\leq n} v_i v_j. \end{aligned} \tag{8.23}$$

Step 3.2: Compute $q_{red}(v_{red})$.
We first solve formally the equation $\nabla g(1, 1, \ldots, 1)\ v = 0$, for an arbitrary vector $v = (v_i)_{i=1:n} \in \mathbb{R}^n$. From (8.21), we find that

$$\nabla g(1, 1, \ldots, 1) \;=\; (1, 1, \ldots, 1).$$

Then, $\nabla g(1, 1, \ldots, 1)\ v = \sum_{i=1}^{n} v_i = 0$, and therefore $v_1 = -\sum_{i=2}^{n} v_i$.

Let $v_{red} = \begin{pmatrix} v_2 \\ \vdots \\ v_n \end{pmatrix}$. We substitute $(-\sum_{i=2}^{n} v_i)$ for v_1 in (8.23), to obtain the reduced quadratic form $q_{red}(v_{red})$, i.e.,

$$\begin{aligned} q_{red}(v_{red}) \;&=\; 2 \left(\sum_{i=2}^{n} v_i \right)^2 \;-\; 2 \sum_{2\leq i<j\leq n} v_i v_j \\ &=\; 2 \left(\sum_{i=2}^{n} v_i^2 \;+\; 2 \sum_{1\leq i<j\leq n} v_i v_j \right) \;-\; 2 \sum_{2\leq i<j\leq n} v_i v_j \\ &=\; 2 \sum_{i=2}^{n} v_i^2 \;+\; 2 \sum_{2\leq i<j\leq n} v_i v_j \\ &=\; \sum_{i=2}^{n} v_i^2 \;+\; \left(\sum_{i=2}^{n} v_i \right)^2. \end{aligned}$$

Then, $q_{red}(v_{red}) > 0$ for any $v_{red} \neq (0, 0, \ldots, 0) \in \mathbb{R}^{n-1}$, and therefore q_{red} is a positive definite quadratic form.
Step 4: From Theorem 8.2, we conclude that $x_0 = (1, 1, \ldots, 1)$ is a constrained minimum point for the function $f(x) = \sum_{i=1}^{n} x_i$ subject to the constraint that

$\prod_{i=1}^{n} x_i = 1$. In other words,

$$\sum_{i=1}^{n} x_i \geq n, \quad \forall \, x_i > 0, \; i = 1 : n, \quad \text{with} \quad \prod_{i=1}^{n} x_i = 1.$$

This is the same as (8.20), which is what we wanted to prove. □

8.2 Numerical methods for one dimensional nonlinear problems

Many problems from finance (and from other fields) require solving equations of the form

$$f(x) \;=\; 0,$$

where $f(x)$ is not a linear function. In sections 8.5–8.7, we discuss several examples of nonlinear problems arising in mathematical finance:

- Computing the yield of a bond;
- Computing the implied volatility;
- Using bootstrapping to find zero rate curves from bond prices.

We present three methods for solving one dimensional nonlinear problems: the bisection method, the secant method, and Newton's method.

8.2.1 Bisection Method

The simplest method for solving one dimensional nonlinear problems is the bisection method. Let $f : [a, b] \to \mathbb{R}$ be a continuous function such that $f(a)$ and $f(b)$ have different signs. From the intermediate value theorem, it follows that the function $f(x)$ is equal to 0 for at least one point x in the interval $[a, b]$. (There might be more points where $f(x) = 0$, but the bisection method will find only one such zero.)

The idea of the bisection method is to divide the interval $[a, b]$ into two equal intervals $[a, c]$ and $[c, b]$, with $c = \frac{a+b}{2}$. Since $f(a)$ and $f(b)$ have different signs, either $f(a)$ and $f(c)$ have different signs, or $f(c)$ and $f(b)$ have different signs (unless $f(c) = 0$, which means that a solution for the problem $f(x) = 0$ has already been found).

If $f(c)$ and $f(a)$ have different signs, we repeat the algorithm for the smaller interval $[a, c]$, called the active interval, and discard the interval $[c, b]$.

If $f(c)$ and $f(b)$ have different signs, we repeat the algorithm for the active interval $[c, b]$, while discarding the interval $[a, c]$.

The bisection method algorithm is stopped, and convergence to a solution of $f(x) = 0$ is declared, when both the value of the function at the midpoint of the active interval and the size of the active interval are small enough; see the pseudocode from Table 8.1 for more details. Possible choices for the tolerance factors are tol_int $= 10^{-6}$ and tol_approx $= 10^{-9}$.

Table 8.1: Pseudocode for the Bisection Method

```
Input:
a, b = left and right endpoints of the initial interval
f(x) = given function
tol_int = largest admissible size of active interval when solution is found
tol_approx = largest admissible value of |f(x)| when solution is found

Output:
x_M = approximate solution for f(x) = 0

x_L = a; x_R = b
while ( max(|f(x_L)|, |f(x_R)|) > tol_approx ) or ( x_R - x_L > tol_int )
  x_M = (x_L + x_R)/2
  if (f(x_L)f(x_M) < 0)
     x_R = x_M          // active interval [x_L, x_M]
  else
     x_L = x_M;         // active interval [x_M, x_R]
  end
end
```

At each step of the bisection method iteration, the size of the active interval is halved. Therefore, the bisection method, while slow to converge in practice, is always convergent.

Theorem 8.3. *If $f : [a, b] \to \mathbb{R}$ is a continuous function and $f(a)$ and $f(b)$ have opposite signs, then the bisection method applied on the interval $[a, b]$ converges to a solution of $f(x) = 0$.*

Rather than giving a formal proof, we present the intuition behind the result of Theorem 8.3. At each time step, the size of the active interval is halved. Thus, after n steps, $x_R - x_L = (b - a)/2^n$, and therefore one of the conditions for stopping the bisection method, $x_R - x_L \leq$ tol_int, will be satisfied for n large enough.

We assume, for simplicity, that the function $f(x)$ is differentiable and that $f'(x)$ is continuous on $[a, b]$. Then, there exists a constant $C > 0$ such that

$$|f'(x)| \leq C, \quad \forall\, x \in [a, b]. \tag{8.24}$$

From the Mean Value Theorem, we know that for any two points x_M and x_L there exists a point $\xi \in (x_L, x_R)$ such that

$$\frac{f(x_R) - f(x_L)}{x_R - x_L} = f'(\xi). \tag{8.25}$$

Note that $x_L, x_R \in [a, b]$. From (8.24) and (8.25) it follows that

$$|f(x_R) - f(x_L)| \leq C\,|x_R - x_L| = C\,\frac{b-a}{2^n}.$$

Since $f(x_R)$ and $f(x_L)$ have different signs, it is easy to see that $|f(x_R) - f(x_L)| = |f(x_R)| + |f(x_L)|$, and therefore

$$\max(|f(x_L)|, |f(x_R)|) \leq |f(x_R) - f(x_L)| \leq C\,\frac{b-a}{2^n}.$$

Then, for n large enough, $\max(|f(x_L)|, |f(x_R)|) <$ tol_approx, and the other stopping condition for the bisection method will be satisfied.

We conclude that, for n large enough, both conditions required to declare that a solution for $f(x) = 0$ is found will be satisfied, and the bisection method will converge.

Example: Use the bisection method on the interval $[-2, 3]$ to find a zero of the function

$$f(x) = x^4 - 5x^2 + 4 - \frac{1}{1 + e^{x^3}}.$$

Answer: To make sure that $f(-2)$ and $f(3)$ have different signs, we first compute $f(-2) = -0.9997$ and $f(3) = 40$. We use the bisection method on the interval $[-2, 3]$ with tol_int $= 10^{-6}$ and tol_approx $= 10^{-9}$. After 33 iterations, the solution -0.889642 is found. Note that $2.000028 \in [-2, 3]$ is another zero for $f(x)$. $\square$

8.2.2 Newton's Method

Newton's method is the most commonly used method for solving nonlinear equations. While the one–dimensional version of Newton's method is easy to explain, its convergence properties (or lack thereof) are somewhat subtle. Unlike the bisection method, Newton's method can easily be extended to N–dimensional problems; see section 8.3.1 for more details.

One way to derive the recursion formula for Newton's method for one–dimensional problems is as follows:

Let x_k be the approximation of the exact solution x^* of $f(x) = 0$ computed after k iterations. The next approximation point x_{k+1} is the x-intercept of the tangent line to the graph of $f(x)$ at the point $(x_k, f(x_k))$.

Mathematically, this is equivalent[4] to the following recursion formula:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad \forall\, k \geq 0. \tag{8.27}$$

A more insightful way of deriving (8.27), which can easily be extended to N-dimensional problems, is to use the Taylor expansion of the function $f(x)$ around the point x_k. From (5.11) for $a = x_k$, we find that

$$f(x) \approx f(x_k) + (x - x_k)f'(x_k). \tag{8.28}$$

Let $x = x_{k+1}$ in (8.28). Then,

$$f(x_{k+1}) \approx f(x_k) + (x_{k+1} - x_k)f'(x_k). \tag{8.29}$$

By approximating $f(x_{k+1})$ by 0 (which is supposed to happen in the limit, if x_{k+1} converges to a solution for $f(x) = 0$), and replacing the approximation sign by an equality sign in (8.29), we find that

$$0 = f(x_k) + (x_{k+1} - x_k)f'(x_k). \tag{8.30}$$

The recursion formula (8.27) is obtained by solving for x_{k+1} in (8.30).

Unlike the bisection method, Newton's method does not necessarily converge for any function $f(x)$ and any initial guess x_0. To obtain convergence in Newton's method, and, in particular, to obtain *fast* convergence, a good choice of the initial approximation x_0 is important.

In general, for financial applications, we have a good idea about what the relevant zeros of the nonlinear function $f(x)$ are. For example, the yield of a bond is positive, usually small, and expressed in percentage points. Therefore, $x_0 = 0.1$, corresponding to an yield of ten percent, is usually a good initial approximation for problems requiring to find the yield of a bond. On the other hand, $x_0 = -10$ may not result in a convergent Newton's method for such problems. Similarly, when computing implied volatility, a good initial

[4]The tangent line to the graph of $f(x)$ passing through the point $(x_k, f(x_k))$ is

$$y - f(x_k) = f'(x_k)\,(x - x_k). \tag{8.26}$$

The x-intercept of this line is found by setting $y = 0$ in (8.26) and solving for x. The solution, denoted by x_{k+1}, is given by (8.27).

choice could be $x_0 = 0.2$, corresponding to 20% volatility, while $x_0 = 10$ or $x_0 = -1$ could be bad choices.

Newton's method is stopped, and convergence to a solution to the problem $f(x) = 0$ is declared, when the following two conditions are satisfied:

$$|f(x_{new})| \leq \text{tol\_approx} \quad \text{and} \quad |x_{new} - x_{old}| \leq \text{tol\_consec}, \tag{8.31}$$

where x_{new} is the most recent value generated by Newton's method and x_{old} is the value previously computed by the algorithm; see the pseudocode from Table 8.2 for more details. Possible choices for the tolerance factors are tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$.

Table 8.2: Pseudocode for Newton's Method

Input:
x_0 = initial guess
$f(x)$ = given function
tol_approx = largest admissible value of $|f(x)|$ when solution is found
tol_consec = largest admissible distance between
two consecutive approximations when solution is found

Output:
x_{new} = approximate solution for $f(x) = 0$

$x_{new} = x_0$; $x_{old} = x_0 - 1$
while ($|f(x_{new})| >$ tol_approx) or ($|x_{new} - x_{old}| >$ tol_consec)
$x_{old} = x_{new}$
$x_{new} = x_{old} - \frac{f(x_{old})}{f'(x_{old})}$
end

If the conditions outlined in Theorem 8.4 below are satisfied, Newton's method is convergent. More importantly, when Newton's method converges, it does so quadratically (i.e., very quickly).

Theorem 8.4. *Let x^* be a solution of $f(x) = 0$. Assume that the function $f(x)$ is twice differentiable with $f''(x)$ continuous. If $f'(x^*) \neq 0$ and if x_0 is close enough to x^*, then Newton's method converges quadratically, i.e., there exists $M > 0$ and n_M a positive integer such that*

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} < M, \quad \forall\, k \geq n_M. \tag{8.32}$$

We provide the intuition behind this result, without giving a complete proof. Since x^* denotes a solution of $f(x) = 0$, it follows that $f(x^*) = 0$.

Then, the Newton's method recursion (8.27) can be written as

$$x_{k+1} - x^* = x_k - x^* - \frac{f(x_k) - f(x^*)}{f'(x_k)}. \tag{8.33}$$

From the derivative formula for the Taylor approximation error (5.3) for $n = 1$, we know that there exists c between a and x such that

$$f(x) - P_1(x) = \frac{(x-a)^2}{2} f''(c),$$

where $P_1(x) = f(a) + (x-a)f'(a)$. Then,

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2} f''(c). \tag{8.34}$$

For $x = x^*$ and $a = x_k$, it follows from (8.34) that there exists c_k between x^* and x_k such that

$$f(x^*) = f(x_k) + (x^* - x_k)f'(x_k) + \frac{(x^* - x_k)^2}{2} f''(c_k). \tag{8.35}$$

From (8.35), we find that

$$\frac{f(x_k) - f(x^*)}{f'(x_k)} = (x_k - x^*) - (x^* - x_k)^2 \frac{f''(c_k)}{2f'(x_k)}.$$

Then, (8.33) can be written as

$$x_{k+1} - x^* = (x^* - x_k)^2 \frac{f''(c_k)}{2f'(x_k)},$$

and we conclude that

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \left| \frac{f''(c_k)}{2f'(x_k)} \right|. \tag{8.36}$$

The rigorous proof of how (8.32) follows from (8.36) is simple, but technical and beyond our scope. We only point out that, since $f''(x)$ and $f'(x)$ are continuous functions and $f'(x^*) \neq 0$, it follows that, if x_k is close to x^*, then $\left|\frac{f''(c_k)}{2f'(x_k)}\right|$ is close to $\left|\frac{f''(x^*)}{2f'(x^*)}\right| < \infty$. Therefore, the term $\left|\frac{f''(c_k)}{2f'(x_k)}\right|$ is uniformly bounded if x_k is close enough to x^*.

Example: Let $f(x) = x^4 - 5x^2 + 4 - \frac{1}{1+e^{x^3}}$. Use Newton's Method to solve $f(x) = 0$.

Answer: We use the algorithm from Table 8.2 with tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$. The problem $f(x) = 0$ has four solutions, which can be found for the following different initial guesses:

for $x_0 = -3$,	the solution	-2.074304	is found after	6	iterations;
for $x_0 = -0.5$,	the solution	-0.889642	is found after	4	iterations;
for $x_0 = 0.5$,	the solution	0.950748	is found after	4	iterations;
for $x_0 = 3$,	the solution	2.000028	is found after	7	iterations.

Note that

$$f'(x) \;=\; 4x^3 - 10x + \frac{3x^2 e^{x^3}}{(1+e^{x^3})^2}. \tag{8.37}$$

Since $f'(0) = 0$ while $f(0) = 3$, it is clear that $x_0 = 0$ would be a poor choice of initial guess, since x_1 given by the recursion (8.27) is not well defined.

If x_0 is close to 0 and positive, than $f'(x_0)$ will also be small and positive, while $f(x_0)$ will be close to 3. Since $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$, we find that x_1 is a much larger positive number. For example, if $x_0 = 0.001$, then $x_1 = 350.0269$. Then, evaluating $f'(x_1)$ using the formula (8.37) will require evaluating $e^{x_1^3}$, which is too large to compute. The iteration will stop in two steps, without converging to a solution. However, $f'(x)$ can also be written as

$$f'(x) \;=\; 4x^3 - 10x + \frac{3x^2 e^{-x^3}}{(1+e^{-x^3})^2}, \tag{8.38}$$

and $f'(x_1)$ can be evaluated for $x_1 = 350.0269$ using (8.38). The solution $x^* = 2.000028$ of $f(x) = 0$ is found in 17 iterations.

It is interesting to evaluate the approximation errors $|x_k - x^*|$ for this last example, i.e., for $x_0 = 0.001$, to see the effect of quadratic convergence:

k	x_k	$\|x_k - x^*\|$
1	35.0326705097	33.032642574
2	26.2923564816	24.292328546
3	19.7430695888	17.743041654
4	14.8390324350	12.839004500
5	11.1715671018	9.1715391666
6	8.4350326120	6.4350046768
7	6.4013415978	4.4013136625
8	4.9009257304	2.9008977951
9	3.8085548302	1.8085268949
10	3.0328229635	1.0327950283
11	2.5083862594	0.5083583242
12	2.1895909404	0.1895630051

k	x_k	$\|x_k - x^*\|$
13	2.0396105458	0.0395826105
14	2.0022975069	0.0022695717
15	2.0000360358	0.0000081005
16	2.0000279353	0.0000000001
17	2.0000279352	0

Recall the quadratic convergence estimate (8.32), i.e.,

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} \;<\; M.$$

For this estimate to be useful, we need $|x_k - x^*| < 1$. Only after this condition is satisfied, convergence will be accelerated.

Indeed, we note that it takes 11 iterations to bring the approximation error $|x_k - x^*|$ below 1, but in the next 6 iterations nine decimal digits are found. The following estimates[5] confirm the quadratic convergence of Newton's method:

$$\frac{|x_{12} - x^*|}{|x_{11} - x^*|^2} = 0.7335; \quad \frac{|x_{13} - x^*|}{|x_{12} - x^*|^2} = 1.1015; \quad \frac{|x_{14} - x^*|}{|x_{13} - x^*|^2} = 1.4486;$$

$$\frac{|x_{15} - x^*|}{|x_{14} - x^*|^2} = 1.5727; \quad \frac{|x_{16} - x^*|}{|x_{15} - x^*|^2} = 1.5808 \quad \Box$$

Finally, note that sensible initial guesses such as 25% volatility and 10% yield for computing implied volatility and bond yields will almost always be within 1 of the exact solution. This is why Newton's method converges very quickly for these problems; see sections 8.5 and 8.6 for details.

8.2.3 Secant Method

For some problems, it is either not possible or not practical to compute $f'(x)$ exactly. For example, in many instances the function $f(x)$ might be given by a numerical routine and a closed formula for $f(x)$, and therefore for $f'(x)$, may not be known. For such problems, Newton's method cannot be used. Instead, a first order finite difference approximation is used to estimate the term $f'(x_k)$ from (8.27), i.e.,

$$f'(x_k) \;\approx\; \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}; \tag{8.39}$$

cf., e.g., (6.5) for $a = x_k$, $h = x_k - x_{k-1}$, and $a - h = x_{k-1}$. By substituting the approximation (8.39) into the Newton's method recursion (8.27), we obtain

[5] Recall that x^* is chosen to be equal to x_{17} when the algorithm is stopped.

the following recursion for the secant method:

$$x_{k+1} = x_k - \frac{(x_k - x_{k-1})f(x_k)}{f(x_k) - f(x_{k-1})}, \quad \forall\, k \geq 0. \tag{8.40}$$

Two approximate guesses x_{-1} and x_0 (with $f(x_{-1}) \neq f(x_0)$) must be made to initialize the secant method. As before, it is important to make guesses that are reasonable given the financial context in which the nonlinear problem occured. The closer x_{-1} and x_0 are to each other, the better the first step of the secant method will approximate the first step of Newton's method with initial guess x_0. The stopping criterion for the secant method is the same as for Newton's method; see (8.31) and Table 8.3 for more details.

Table 8.3: Pseudocode for the Secant Method

```
Input:
x_{-1}, x_0 = initial guesses
f(x) = given function
tol_approx = largest admissible value of |f(x)| when solution is found
tol_consec = largest admissible distance between
        two consecutive approximations when solution is found

Output:
x_new = approximate solution for f(x) = 0

x_new = x_0; x_old = x_{-1}
while ( |f(x_new)| > tol_approx ) or ( |x_new - x_old| > tol_consec )
  x_oldest = x_old
  x_old = x_new
  x_new = x_old - f(x_old) (x_old - x_oldest)/(f(x_old) - f(x_oldest))
end
```

As expected, since it is derived using Newton's method as a starting point, the secant method does not necessarily converge for every function $f(x)$, nor for every initial guesses x_{-1} and x_0. Also, it can only be proved that the secant method is linearly convergent, and not quadratically convergent, as was the case for Newton's method. Therefore, the secant method is usually slower than Newton's method.

Example: Use the secant method to find a zero of the function

$$f(x) = x^4 - 5x^2 + 4 - \frac{1}{1+e^{x^3}}.$$

Answer: We choose $x_0 = -3$ and $x_{-1} = x_0 - 0.01$. The secant method with tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$ converges after 8 iterations to the solution -2.074304 of $f(x) = 0$.

For comparison purposes, note that Newton's method with initial guess $x_0 = -3$ converges in 6 iterations to the same solution -2.074304 of $f(x) = 0$:

	Secant Method	Newton's Method
x_1	-2.503129020	-2.5
x_3	-2.309651708	-2.211666639
x_3	-2.159035029	-2.094956590
x_4	-2.095347041	-2.074874887
x_5	-2.076550029	-2.074304856
x_6	-2.074368712	-2.074304403
x_7	-2.074304603	
x_8	-2.074304403	

8.3 Numerical methods for N–dimensional nonlinear problems

For N dimensional problems there are no practical analogs of the bisection method. However, Newton's method can easily be extended to N dimensional problems. An approximate Newton's method similar to the secant method and based on finite difference–type approximations also exists.

8.3.1 The N–dimensional Newton's Method

Let $F : \mathbb{R}^N \to \mathbb{R}^N$ given by $F(x) = (F_i(x))_{i=1:N}$, where $F_i : \mathbb{R}^N \to \mathbb{R}$. Assume that all the partial order derivatives of the functions $F_i(x)$, $i = 1 : N$, are continuous. We want to solve the nonlinear N-dimensional problem

$$F(x) = 0.$$

Recall from (1.38) that the gradient $DF(x)$ of $F(x)$ is an $N \times N$ matrix:

$$DF(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \dots & \frac{\partial F_1}{\partial x_N}(x) \\ \frac{\partial F_2}{\partial x_1}(x) & \frac{\partial F_2}{\partial x_2}(x) & \dots & \frac{\partial F_2}{\partial x_N}(x) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_N}{\partial x_1}(x) & \frac{\partial F_N}{\partial x_2}(x) & \dots & \frac{\partial F_N}{\partial x_N}(x) \end{pmatrix}.$$

The recursion for Newton's method follows from the linear Taylor expansion of $F(x)$, i.e.,

$$F(x) \approx F(x_k) + DF(x_k)(x - x_k); \tag{8.41}$$

cf. (5.34) for $a = x_k$. Here, $DF(x_k)(x - x_k)$ is a matrix–vector multiplication. Let $x = x_{k+1}$ in (8.41). Approximating $F(x_{k+1})$ by 0 (which happens in the limit, if convergence to a solution for $F(x) = 0$ is achieved), we find that

$$0 \approx F(x_k) + DF(x_k)(x_{k+1} - x_k). \tag{8.42}$$

Changing (8.42) into an equality and solving for x_{k+1}, we obtain the following recursion for Newton's method for N dimensional problems:

$$x_{k+1} = x_k - (DF(x_k))^{-1}F(x_k), \quad \forall\, k \geq 0. \tag{8.43}$$

At each step, the vector $(DF(x_k))^{-1}F(x_k)$ must be computed. In practice, the inverse matrix $(DF(x_k))^{-1}$ is never explicitely computed, since this would be very expensive computationally. Instead, we note that computing the vector $v_k = (DF(x_k))^{-1}F(x_k)$ is equivalent to solving the linear system

$$DF(x_k)v_k = F(x_k).$$

This is done using numerical linear algebra methods, e.g., by computing the LU decomposition factors of the matrix $DF(x_k)$ and then doing a forward and a backward substitution.

It is not our goal here to discuss such methods; see [27] for details on numerical linear algebra methods. We subsequently assume that a routine for solving linear systems called solve_linear_system exists such that, given a nonsingular square matrix A and a vector b, the vector

$$x = \text{solve\_linear\_system}(A, b)$$

is the unique solution of the linear system $Ax = b$.

The vector $v_k = (DF(x_k))^{-1}F(x_k)$ can then be computed as

$$v_k = \text{solve\_linear\_system}(DF(x_k), F(x_k)),$$

and recursion (8.43) can be written as

$$x_{k+1} = x_k - \text{solve\_linear\_system}(DF(x_k), F(x_k)), \quad \forall\, k \geq 0.$$

The N–dimensional Newton's method iteration is stopped and convergence to a solution to the problem $F(x) = 0$ is declared when the following two conditions are satisfied:

$$||F(x_{new})|| \leq \text{tol\_approx} \quad \text{and} \quad ||x_{new} - x_{old}|| \leq \text{tol\_consec}, \tag{8.44}$$

where x_{new} is the most recent value generated by Newton's method and x_{old} is the value previously computed by the algorithm. Here, $||\cdot||$ represents

the Euclidean norm[6]. Possible choices for the tolerance factors are tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$; see the pseudocode from Table 8.2 for more details.

Table 8.4: Pseudocode for the N-dimensional Newton's Method

```
Input:
x_0 = initial guess
F(x) = given function
tol_approx = largest admissible value of ||F(x)|| when solution is found
tol_consec = largest admissible distance between
        two consecutive approximations when solution is found

Output:
x_new = approximate solution for f(x) = 0

x_new = x_0; x_old = x_0 - 1
while ( ||F(x_new)|| > tol_approx ) or ( ||x_new - x_old|| > tol_consec )
  x_old = x_new
  compute DF(x_old)
  x_new = x_old - solve_linear_system(DF(x_old), F(x_old))
end
```

As was the case for the one–dimensional version, the N–dimensional Newton's method converges quadratically if certain conditions are satisfied.

Theorem 8.5. *Let x^* be a solution of $F(x) = 0$, where $F(x)$ is a function with continuous second order partial derivatives. If $DF(x^*)$ is a nonsingular matrix, and if x_0 is close enough to x^*, then Newton's method converges quadratically, i.e., there exists $M > 0$ and n_M a positive integer such that*

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||^2} < M, \quad \forall\, k \geq n_M.$$

Example: Use Newton's method to solve $F(x) = 0$, for

$$F(x) = \begin{pmatrix} x_1^3 + 2x_1x_2 + x_3^2 - x_2x_3 + 9 \\ 2x_1^2 + 2x_1x_2^2 + x_2^3x_3^2 - x_2^2x_3 - 2 \\ x_1x_2x_3 + x_1^3 - x_3^2 - x_1x_2^2 - 4 \end{pmatrix}$$

[6] If $v = (v_i)_{i=1:N}$ is a vector in $\mathbb{R}^N$, then

$$||v|| = \left(\sum_{i=1}^{N} |v_i|^2\right)^{1/2}.$$

Answer: Note that

$$DF(x) = \begin{pmatrix} 3x_1^2 + 2x_2 & 2x_1 - x_3 & 2x_3 - x_2 \\ 4x_1 + 2x_2^2 & 4x_1x_2 + 3x_2^2x_3^2 - 2x_2x_3 & 2x_2^3x_3 - x_2^2 \\ x_2x_3 + 3x_1^2 - x_2^2 & x_1x_3 - 2x_1x_2 & x_1x_2 - 2x_3 \end{pmatrix}.$$

We use the algorithm from Table 8.4 with tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$. For the initial guess

$$x_0 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \text{the solution } x^* = \begin{pmatrix} -1.690550759854953 \\ 1.983107242868416 \\ -0.884558078475291 \end{pmatrix}$$

is found after 9 iterations.

For the initial guess $x_0 = (2\ 2\ 2)^t$, the solution $x^* = (-1\ 3\ 1)^t$ is found after 40 iterations. □

8.3.2 The Approximate Newton's Method

In many instances, it is not possible (or efficient) to find a closed formula for the matrix $DF(x)$ which is needed for Newton's method; cf. (8.43). In these cases, finite difference approximations can be used to estimate each entry of $DF(x)$. The resulting method is called the Approximate Newton's Method.

The entry of $DF(x)$ on the position (i,j), i.e., $\frac{\partial F_i}{\partial x_j}(x)$, is estimated using forward finite differences approximations (6.2) as

$$\frac{\partial F_i}{\partial x_j}(x) \approx \Delta_j F_i(x) = \frac{F_i(x + he_j) - F_i(x)}{h}, \tag{8.45}$$

where h is a small number and e_j is a vector with all entries equal to 0 with the exception of the j-th entry, which is equal to 1, i.e., $e_j(k) = 0$, for all $k \neq j$ and $e_j(j) = 1$. Note that if $x = (x_i)_{i=1:N}$, then

$$x + he_j = \begin{pmatrix} x_1 \\ \vdots \\ x_{j-1} \\ x_j + h \\ x_{j+1} \\ \vdots \\ x_N \end{pmatrix}.$$

Let

$$\Delta F(x) = \begin{pmatrix} \Delta_1 F_1(x) & \Delta_2 F_1(x) & \ldots & \Delta_N F_1(x) \\ \Delta_1 F_2(x) & \Delta_2 F_2(x) & \ldots & \Delta_N F_2(x) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta_1 F_N(x) & \Delta_2 F_N(x) & \ldots & \Delta_N F_N(x) \end{pmatrix}.$$

The recursion formula for the approximate Newton's method is obtained by replacing $DF(x)$ with $\Delta F(x)$ in the recursion (8.43) for the N–dimensional Newton's method, i.e.,

$$x_{k+1} = x_k - (\Delta F(x_k))^{-1} F(x_k), \tag{8.46}$$

The stopping criterion and the pseudocode for this method are similar to those for the N–dimensional Newton's method; cf. (8.44) and Table 8.5.

Table 8.5: Pseudocode for the N-dimensional Approximate Newton's Method

```
Input:
x_0 = initial guess
F(x) = given function
h = parameter for the finite difference approximations of
        partial derivatives in the approximate gradient ΔF(x)
tol_approx = largest admissible value of ||F(x)|| when solution is found
tol_consec = largest admissible distance between
        two consecutive approximations when solution is found

Output:
x_new = approximate solution for f(x) = 0

h = tol_consec; x_new = x_0; x_old = x_0 - 1
while ( ||F(x_new)|| > tol_approx ) or ( ||x_new - x_old|| > tol_consec )
  x_old = x_new
  compute ΔF(x_old)        // use parameter h
  x_new = x_old - solve_linear_system(ΔF(x_old), F(x_old))
end
```

Note that a more precise estimate of the entries of $DF(x)$ can be obtained using central finite differences (6.6) to approximate $\frac{\partial F_i}{\partial x_j}(x)$ as follows:

$$\frac{\partial F_i}{\partial x_j}(x) \approx \frac{F_i(x + he_j) - F_i(x - he_j)}{2h}.$$

The drawback of this method is that the function $F(x)$ has to be evaluated twice as many times than if the forward difference approximation (8.45) is used. This might be computationally expensive and undermine the potential savings in terms of the number of iterations required until convergence generated by having a more precise approximation of the gradient $DF(x)$.

Example: Use the approximate Newton's method to solve $F(x) = 0$, for

$$F(x) = \begin{pmatrix} x_1^3 + 2x_1x_2 + x_3^2 - x_2x_3 + 9 \\ 2x_1^2 + 2x_1x_2^2 + x_2^3x_3^2 - x_2^2x_3 - 2 \\ x_1x_2x_3 + x_1^3 - x_3^2 - x_1x_2^2 - 4 \end{pmatrix}$$

Answer: We use the algorithm from Table 8.4 with tol_consec $= 10^{-6}$ and tol_approx $= 10^{-9}$. The parameter h is chosen to be equal to tol_consec, i.e., $h = 10^{-6}$, and forward finite differences are used for computing $\Delta F(x)$.

For the initial guess

$$x_0 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \text{the solution } x^* = \begin{pmatrix} -1.690550759854953 \\ 1.983107242868416 \\ -0.884558078475291 \end{pmatrix}$$

is found after 9 iterations. This is the same number of iterations as required by Newton's method with the same initial guess.

For the initial guess $x_0 = (2\ 2\ 2)^t$, the solution $x^* = (-1\ 3\ 1)^t$ is found after 58 iterations. Recall from section 8.3 that Newton's method required 40 iterations to converge in this case. □

FINANCIAL APPLICATIONS

Finding optimal investment portfolios.

Computing the yield of a bond.

Computing implied volatility from the Black–Scholes model.

The bootstrapping method for finding zero rate curves.

8.4 Finding optimal investment portfolios

Consider a portfolio with investments in n assets. Let w_i be the proportion of the portfolio invested in asset i. (In other words, asset i has weight w_i in the portfolio.) Then

$$\sum_{i=1}^{n} w_i = 1. \tag{8.47}$$

We assume it is possible to take arbitrarily large short positions in any of the assets. Therefore, the weights w_i are not required to be positive[7].

Let R_i be the rate of return (over a fixed period of time) of asset i, and let $\mu_i = E[R_i]$ and $\sigma_i^2 = \mathrm{var}(R_i)$ be the expected value and variance of R_i,

[7] If short selling is not allowed, then all assets must have positive weights, i.e., $w_i \geq 0$, $i = 1 : n$. These inequality constraints make the portfolio optimization problem a quadratic programming problem which cannot be solved using Lagrange multipliers.

respectively, for $i = 1 : n$. Let $\rho_{i,j}$ be the correlation between the rates of return R_i and R_j, for $1 \leq i < j \leq n$.

The rate of return R of the portfolio is

$$R = \sum_{i=1}^{n} w_i R_i. \tag{8.48}$$

From Lemma 3.9, it follows that

$$E[R] = \sum_{i=1}^{n} w_i \mu_i; \tag{8.49}$$

$$\mathrm{var}(R) = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{1 \leq i < j \leq n} w_i w_j \sigma_i \sigma_j \rho_{i,j}. \tag{8.50}$$

We note that the expected values and variances of the rates of return of the n assets, as well as the correlations between any two assets, can be estimated, e.g., by using historical data.

A natural question to ask is how to make the portfolio selection efficient[8]? In other words, how do you choose the weights w_i, $i = 1 : n$, to obtain a portfolio with smallest variance of the rate of return, given a fixed expected rate of return, or to obtain a portfolio with the highest rate of return given a fixed variance?

Given μ_P, find w_i, $i = 1 : n$, with $E[R] = \mu_P$, such that var(R) is minimal.

Given σ_P, find w_i, $i = 1 : n$, with var$(R) = \sigma_P^2$, such that $E[R]$ is maximal.

These are constrained optimization problems and can be solved using Lagrange multipliers.

Rather than discuss the general case of n assets, we provide more details for a particular example:

Find a portfolio invested in $n = 4$ assets with expected rate of return equal to μ_P and minimum variance of the return. Assume that the return of one of the assets is independent of the returns of the other three assets. Also, assume that not all assets have the same expected rate of return.

For notation purposes, assume that asset 4 is the asset with uncorrelated return, i.e., $\rho_{i,4} = 0$, for $i = 1 : 3$. Let w_i be the weight of asset i in the

[8] Finding efficient portfolios is one of the fundamental problems answered by the modern portfolio theory of Markowitz and Sharpe; see Markowitz [17] and Sharpe [25] for seminal papers. Of all the efficient portfolios, the portfolio with the highest Sharpe ratio $\frac{E[R]-r_f}{\sigma(R)}$, i.e., the expected return above the risk free rate r_f normalized by the standard deviation $\sigma(R)$ of the return, is called the market portfolio (or the tangency portfolio) and plays an important role in the Capital Asset Pricing Model (CAPM).

portfolio, for $i = 1 : 4$. From (8.49) and (8.50), it follows that

$$\begin{aligned} E[R] &= w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4; &(8.51)\\ \mathrm{var}(R) &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + w_4^2\sigma_4^2 &(8.52)\\ &\quad +2\,(w_1w_2\sigma_1\sigma_2\rho_{1,2} + w_2w_3\sigma_2\sigma_3\rho_{2,3} + w_1w_3\sigma_1\sigma_3\rho_{1,3}). \end{aligned}$$

Recall from (8.47) that $w_1+w_2+w_3+w_4 = 1$. We are looking for a portfolio with $E[R] = \mu_P$, where μ_P is the given expected rate of return, such that var(R) minimal. Using (8.51), we conclude that the weights $(w_i)_{i=1:4}$ must satisfy the following two constraints:

$$\begin{aligned} w_1 + w_2 + w_3 + w_4 &= 1; &(8.53)\\ w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4 &= \mu_P. &(8.54) \end{aligned}$$

From (8.52–8.54), it follows that this problem can be written as a constrained optimization problem as follows: find w_0 such that

$$\min_{g(w)=0} f(w) \;=\; f(w_0), \qquad (8.55)$$

where $w = (w_i)_{i=1:4}$, and $f : \mathbb{R}^4 \to \mathbb{R}$ and $g : \mathbb{R}^4 \to \mathbb{R}^2$ are defined as

$$\begin{aligned} f(w) &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + w_4^2\sigma_4^2 &(8.56)\\ &\quad +2\,(w_1w_2\sigma_1\sigma_2\rho_{1,2} + w_2w_3\sigma_2\sigma_3\rho_{2,3} + w_1w_3\sigma_1\sigma_3\rho_{1,3});\\ g(w) &= \begin{pmatrix} g_1(w) \\ g_2(w) \end{pmatrix} = \begin{pmatrix} w_1 + w_2 + w_3 + w_4 - 1 \\ w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4 - \mu_P \end{pmatrix}. &(8.57) \end{aligned}$$

We first check that condition (8.9) is satisfied. Note that $n = 4$ and $m = 2$ for this problem. It is easy to see that

$$\nabla g(w) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \end{pmatrix}.$$

Note that $\nabla g(w)$ is a constant matrix, i.e., a constant function of w. Since we assumed that not all assets have the same expected rate of return, the matrix $\nabla g(w)$ has rank 2 for any w, and therefore condition (8.9) is satisfied.

Denote by $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$ a Lagrange multiplier associated to this problem. From (8.2), and using (8.56) and (8.57), we obtain the Lagrangian

$$\begin{aligned} F(w,\lambda) &= f(w) \;+\; \lambda_1\, g_1(w) \;+\; \lambda_2\, g_2(w) &(8.58)\\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + w_4^2\sigma_4^2\\ &\quad +\, 2\,(w_1w_2\sigma_1\sigma_2\rho_{1,2} + w_2w_3\sigma_2\sigma_3\rho_{2,3} + w_1w_3\sigma_1\sigma_3\rho_{1,3})\\ &\quad +\, \lambda_1\,(w_1 + w_2 + w_3 + w_4 - 1)\\ &\quad +\, \lambda_2\,(w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4 - \mu_P). \end{aligned}$$

The gradient $\nabla_{(w,\lambda)} F(w,\lambda)$ is the following (row) vector:

$$\nabla_{(x,\lambda)} F(x,\lambda) = \begin{pmatrix} 2w_1\sigma_1^2 + 2w_2\sigma_1\sigma_2\rho_{1,2} + 2w_3\sigma_1\sigma_3\rho_{1,3} \;+\; \lambda_1 \;+\; \lambda_2\mu_1 \\ 2w_2\sigma_2^2 + 2w_1\sigma_1\sigma_2\rho_{1,2} + 2w_3\sigma_2\sigma_3\rho_{2,3} \;+\; \lambda_1 \;+\; \lambda_2\mu_2 \\ 2w_3\sigma_3^2 + 2w_1\sigma_1\sigma_3\rho_{1,3} + 2w_2\sigma_2\sigma_3\rho_{2,3} \;+\; \lambda_1 \;+\; \lambda_2\mu_3 \\ 2w_4\sigma_4^2 \;+\; \lambda_1 \;+\; \lambda_2\mu_4 \\ w_1 + w_2 + w_3 + w_4 - 1 \\ w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4 - \mu_P \end{pmatrix}^t.$$

To find the critical points of $F(w,\lambda)$, we solve $\nabla_{(w,\lambda)} F(w,\lambda) = 0$, which can be written as a linear system as follows:

$$\begin{pmatrix} 2\sigma_1^2 & 2\sigma_1\sigma_2\rho_{1,2} & 2\sigma_1\sigma_3\rho_{1,3} & 0 & 1 & \mu_1 \\ 2\sigma_1\sigma_2\rho_{1,2} & 2\sigma_2^2 & 2\sigma_2\sigma_3\rho_{2,3} & 0 & 1 & \mu_2 \\ 2\sigma_1\sigma_3\rho_{1,3} & 2\sigma_2\sigma_3\rho_{2,3} & 2\sigma_3^2 & 0 & 1 & \mu_3 \\ 0 & 0 & 0 & 2\sigma_4^2 & 1 & \mu_4 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \mu_P \end{pmatrix}. \qquad (8.59)$$

Assume that the matrix corresponding to the linear system (8.59) is nonsingular[9]. Let (w_0, λ_0) be the unique solution of (8.59). Since condition (8.9) is satisfied, we know from Theorem 8.1 that w_0 is the only possible solution for the constrained optimization problem (8.55). To identify whether w_0 is, indeed, a constrained minimum, we need to construct the reduced quadratic form $q_{red}(v_{red})$ given by (8.14). Let $\lambda_0 = \begin{pmatrix} \lambda_{0,1} \\ \lambda_{0,2} \end{pmatrix}$ and define

$$F_0(w) \;=\; F(w,\lambda_0) \;=\; f(w) \;+\; \lambda_{0,1}\, g_1(w) \;+\; \lambda_{0,2}\, g_2(w).$$

The Hessian $D^2F_0(w)$ is

$$D^2F_0(w) = \begin{pmatrix} 2\sigma_1^2 & 2\sigma_1\sigma_2\rho_{1,2} & 2\sigma_1\sigma_3\rho_{1,3} & 0 \\ 2\sigma_1\sigma_2\rho_{1,2} & 2\sigma_2^2 & 2\sigma_2\sigma_3\rho_{2,3} & 0 \\ 2\sigma_1\sigma_3\rho_{1,3} & 2\sigma_2\sigma_3\rho_{2,3} & 2\sigma_3^2 & 0 \\ 0 & 0 & 0 & 2\sigma_4^2 \end{pmatrix}.$$

Since the Hessian $D^2F_0(w)$ is a constant matrix for any $w \in \mathbb{R}^4$, a shortcut is possible for this particular problem. It is easy to see that $D^2F_0(w)$ is a

[9] The matrix corresponding to the linear system (8.59) is singular if all assets have the same expected rate of return, or if two assets have identical rates of return.

positive definite matrix[10]. Therefore, the associated quadratic form

$$q(v) = v^t \ D_w^2 F(w_0, \lambda_0) \ v$$

is positive definite, and so will be the reduced quadratic form $q_{red}(v_{red})$, regardless of what the linear constraints $\nabla g(x_0) \ v = 0$ are. Then, from Theorem 8.2, we obtain that w_0 is a constrained minimum for $f(w)$ given the constraints $g(w) = 0$. In other words, a portfolio where the weight of asset i in the portfolio is equal to $w_{0,i}$, for $i = 1:4$, is the minimum variance portfolio with rate of return μ_P. □

Example: Find a minimal variance portfolio with 11.5% expected rate of return, if four assets can be traded to set up the portfolio, given the following data on the rates of return of the assets:

$$\begin{array}{llllll} \mu_1 = 0.09; & \sigma_1 = 0.2; & \rho_{1,2} = -0.5; \\ \mu_2 = 0.12; & \sigma_2 = 0.3; & \rho_{2,3} = 0.25; \\ \mu_3 = 0.15; & \sigma_3 = 0.35; & \rho_{1,3} = 0.35; \\ \mu_4 = 0.06; & \sigma_4 = 0.15; & \rho_{i,4} = 0, \ \forall\, i = 1:3. \end{array}$$

Answer: As seen above, the weights for the minimum variance portfolio are the first four entries of the solution to the linear system (8.59) that identifies the critical points of the Lagrangian function of the corresponding constrained optimization problem.

By direct computation, we find that the system (8.59) can be written as

$$\begin{pmatrix} 0.08 & -0.06 & 0.049 & 0 & 1 & 0.09 \\ -0.06 & 0.18 & 0.0525 & 0 & 1 & 0.12 \\ 0.049 & 0.0525 & 0.245 & 0 & 1 & 0.15 \\ 0 & 0 & 0 & 0.045 & 1 & 0.06 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0.09 & 0.12 & 0.15 & 0.06 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0.115 \end{pmatrix}.$$

[10]Note that $D^2F_0(w)$ is the sum of the diagonal matrix

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}$$

and the covariance matrix of the rates of return $(R_i)_{i=1:4}$, which is a positive definite matrix.

The solution to this system is

$$\begin{pmatrix} w_{0,1} \\ w_{0,2} \\ w_{0,3} \\ w_{0,4} \\ \lambda_{0,1} \\ \lambda_{0,2} \end{pmatrix} = \begin{pmatrix} 0.547452 \\ 0.440377 \\ 0.135042 \\ -0.122872 \\ 0.064569 \\ -0.983995 \end{pmatrix}.$$

The asset allocation for a minimal variance portfolio with 11.5% expected rate of return is as follows: 54.75% in asset 1, 44.03% in asset 2, 13.5% in asset 3, while shorting an amount of asset 4 equal to 12.29% of the value of the portfolio. For example, if the value of the portfolio is $1,000,000, then $122,872 of asset 4 is shorted (borrowed and sold for cash) $547,452 is invested in asset 1, $440,377 is invested in asset 2, and $135,042 is invested in asset 3.

The minimal variance portfolio has a standard deviation of the expected rate of return equal to 15.5869%. □

8.5 Computing bond yields

Recall from section 2.7 that a bond is a financial instrument where the holder of the bond receives periodic payments until the maturity of the bond.

Let c_i, $i = 1:n$, be all the cash payments received by the holder of the bond, and let t_i be the time when payment c_i is made. The yield of the bond y is the internal rate of return of the bond, and is related to the value B of the bond by formula (2.56), i.e.,

$$B = \sum_{i=1}^{n} c_i e^{-yt_i}. \tag{8.60}$$

If the price of the bond and all its future cash flows are known, finding the yield y of the bond means solving equation (8.60) for y. We recast the problem of finding the yield of the bond as the nonlinear equation

$$f(x) = 0, \tag{8.61}$$

by setting $x = y$ and defining $f(x)$ as

$$f(x) = \sum_{i=1}^{n} c_i e^{-xt_i} - B. \tag{8.62}$$

We solve problem (8.61) using Newton's method. It is easy to see that

$$f'(x) = -\sum_{i=1}^{n} t_i c_i e^{-x t_i}. \tag{8.63}$$

Using (8.62) and (8.63), we find that the recursion formula (8.27) for solving (8.61) becomes

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k + \frac{\sum_{i=1}^n c_i e^{-x_k t_i} - B}{\sum_{i=1}^n t_i c_i e^{-x_k t_i}}.$$

A good initial guess is a ten percent yield, i.e., $x_0 = 0.1$. The algorithm is stopped when the difference between two consecutive approximations of the yield is less than 10^{-6}; see Table 8.6 for the pseudocode.

Table 8.6: Pseudocode for computing a bond yield

```
Input:
B = the price of the bond
n = number of cash flows
t_cash_flow = cash flow dates (vector of size n)
v_cash_flow = cash flows (vector of size n)
tol = tolerance for declaring convergence of Newton's method

Output:
x_new = the yield of the bond

x_0 = 0.1;            // initial guess 10%
x_new = x_0; x_old = x_0 - 1; tol = 10^(-6)
while (|x_new - x_old| > tol)
   x_old = x_new
   x_new = x_old + (sum_{i=1}^n v_cash_flow(i) exp(-x_old t_cash_flow(i)) - B)
                 / (sum_{i=1}^n t_cash_flow(i) v_cash_flow(i) exp(-x_old t_cash_flow(i)))
end
```

Example: A 34 months semiannual coupon bond with coupon rate 8% is priced at 105. The face value of the bond is assumed to be 100. Compute the yield of the bond.

Answer: For a semiannual bond, one coupon payment is made every six months. Since the final payment will be made in 34 months, when the bond expires, we conclude that there are $n = 6$ coupon payments made in 4, 10, 16, 22, 28, and 34 months, corresponding to

$$\text{t\_cash\_flow} = \left[\frac{4}{12}\ \frac{10}{12}\ \frac{16}{12}\ \frac{22}{12}\ \frac{28}{12}\ \frac{34}{12}\right].$$

Each coupon payment is equal to the face value (100) times the coupon rate (0.08) divided by two since two payments are made every year for a semiannual bond. At maturity, both a coupon payment is made, and the face value is returned to the holder of the bond. Therefore,

$$\text{v\_cash\_flow} = [4\ 4\ 4\ 4\ 4\ 104].$$

With initial guess $x_0 = 0.1$, Newton's method converges in four iterations:

$$\begin{aligned} x_1 &= 0.062726241677766; \\ x_2 &= 0.064497631933507; \\ x_3 &= 0.064501931395564; \\ x_4 &= 0.064501931420810. \end{aligned}$$

We conclude that the yield of the bond is 6.4502%. □

8.6 Implied volatility

For any plain vanilla European call or put option, the maturity and strike of the option are given. Also, when the option is traded, the option price and the spot price of the underlying asset are known. If we assume a lognormal model for the evolution of the underlying asset, then the price of the option is given by the Black–Scholes formulas (3.53–3.56) for $t = 0$, i.e.,

$$C = Se^{-qT}N(d_1) - Ke^{-rT}N(d_2); \tag{8.64}$$

$$P = Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1), \tag{8.65}$$

where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}; \quad d_2 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

We assumed that the underlying asset pays dividends continuously at rate q and that the interest rates are constant and equal to r.

Note that C, P, S, K, and T are known in formulas (8.64) and (8.65). Assume that the dividend yield q and the constant interest rate r can be estimated from market data. The implied volatility is the unique value of the volatility parameter σ in the lognormal model that makes the Black–Scholes value of the option equal to the price the option traded at.

If we look at (8.64) and (8.65) as functions of only one variable, σ, finding the implied volatility requires solving the nonlinear problem

$$f(x) = 0, \tag{8.66}$$

where $x = \sigma$ and

$$f(x) = Se^{-qT}N(d_1(x)) - Ke^{-rT}N(d_2(x)) - C, \qquad (8.67)$$

for the call option, and

$$f(x) = Ke^{-rT}N(-d_2(x)) - Se^{-qT}N(-d_1(x)) - P, \qquad (8.68)$$

for the put option. Here,

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{x^2}{2}\right)T}{x\sqrt{T}}; \quad d_2 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q - \frac{x^2}{2}\right)T}{x\sqrt{T}}.$$

The value of σ thus computed is the implied volatility corresponding to a given option price.

It is interesting to note that computing the implied volatility is a straightforward way of showing that the lognormal assumption, and the Black–Scholes formulas derived based on this assumption, are not correct. At any point in time, several options with different strikes and maturities may be traded. If the lognormal assumption were true, then the implied volatilities corresponding to all these options should be equal. However, this does not happen. Usually, the implied volatility of either deep out of the money or deep in the money options is higher than the implied volatility of at the money options. This phenomenon is called the volatility smile. Another possible pattern for implied volatility is the volatility skew, when, e.g., the implied volatility of deep in the money options is smaller than the implied volatility of at the money options, which in turn is smaller than the implied volatility of deep out of the money options.

We restrict our attention to solving the problem (8.66) corresponding to the call option, i.e., with the function $f(x)$ given by (8.67), using Newton's method. Note that differentiating the function $f(x)$ with respect to x is the same as computing the vega of the call option. From (3.70), we find that

$$f'(x) = \frac{1}{\sqrt{2\pi}} Se^{-qT}\sqrt{T}\exp\left(-\frac{(d_1(x))^2}{2}\right). \qquad (8.69)$$

Since $f'(x) \geq 0$ for any x, it follows that the function $f(x)$ is a strictly increasing function, and therefore the problem (8.66) has at most one solution.

The recursion (8.27) for Newton's method for solving (8.66) is

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \qquad (8.70)$$

where the functions $f(x)$ and $f'(x)$ are given by (8.67) and (8.69), respectively. A good initial guess is 25% volatility, i.e., $x_0 = 0.25$. The algorithm

Table 8.7: Pseudocode for computing implied volatility

```
Input:
C = price of the call option
S = spot price of the underlying asset
K = strike price of the option
T = maturity of the option
q = dividend yield of the underlying asset
r = risk-free constant rate
tol = tolerance for declaring convergence of Newton's method
f_BS(x) = Black–Scholes value of a call; here, x = volatility
vega_BS(x) = vega of the call; x = volatility

Output:
x_new = implied volatility

x_0 = 0.25;                // initial guess: 25% volatility
x_new = x_0; x_old = x_0 - 1; tol = 10^(-6)
while (|x_new - x_old| > tol)
  x_old = x_new
  x_new = x_new - (f_BS(x_new) - C) / vega_BS(x_new)
end
```

is stopped when two consecutive approximations in Newton's method are within 10^{-6} of each other; see the pseudocode from Table 8.7.

Example: The price of a one year call option with strike 20 on a non–dividend paying asset with spot price 25 is \$7. Assuming that the risk–free interest rate is constant at 5% over the life of the option, find the implied volatility.

Answer: Note that $C = 7$, $S = 25$, $K = 20$, $T = 1$, and $r = 0.05$. We apply Newton's method (8.70) to solve the nonlinear equation (8.66), with $f(x)$ and $f'(x)$ given by (8.67) and (8.69), respectively.

With initial guess $x_0 = 0.25$, the iteration converges in four steps:

$$\begin{aligned} x_1 &= 0.386109169107613; \\ x_2 &= 0.363400625387253; \\ x_3 &= 0.363063304424297; \\ x_4 &= 0.363063222956668. \end{aligned}$$

The implied volatility is 36.3063%. □

8.7 Bootstrapping for finding zero rate curves

Recall from section 2.6 that the zero rate $r(0,t)$ between time 0 and time t is the rate of return of a cash deposit made at time 0 and maturing at time t. (We assume that interest is compounded continuously.) Also recall from (2.53) that the price B of a bond with future cash flows c_i to be paid to the holder of the bond on dates t_i, $i = 1:n$, is given by

$$B = \sum_{i=1}^{n} c_i e^{-r(0,t_i)t_i}. \tag{8.71}$$

If the prices of several different bonds are known, then we are provided with information on the zero rates corresponding to each date when a bond makes a payment. Since this is a discrete and finite set of data, the general form of the zero rate curve $r(0,t)$ for all times t smaller than the largest maturity of any of the bonds cannot be determined uniquely.

The bootstrapping method provides a simplified framework in which the zero rate curve can be uniquely determined from a given set of bond prices (assuming zero credit risk, i.e., that the bonds never default), by assuming that the zero rate $r(0,t)$ is linear between any two consecutive bond maturities.

For clarity purposes, rather than formalizing the bootstrapping method using abstract notations, we present a concrete example:

Example: The prices and coupon rates of four semiannual coupon bonds with face value 100 are as follows:

Maturity	Coupon Rate	Price
6 months	0	99
1 year	4	102
2 years	4	103.5
5 years	4	109

We use bootstrapping to fit a zero rate interest curve $r(0,t)$, $0 \le t \le 5$ to the prices of these bonds, by assuming that the zero rate curve in linear between every two consecutive maturities of the bonds. To do so, we use formula (2.55) for the price of a semiannual coupon bond, i.e.,

$$B = \sum_{i=1}^{n-1} \frac{C}{2}\, 100\, e^{-r(0,t_i)t_i} + \left(100 + \frac{C}{2}\, 100\right) e^{-r(0,T)T}, \tag{8.72}$$

where C is the coupon rate, T is the maturity of the bond, and t_i, $i = 1:(n-1)$, are all the coupon dates before maturity.

We look at the bonds sequentially, from the bond with the shortest maturity to the bond with the longest maturity.

The six months semiannual coupon bond with coupon rate equal to 0 is, in fact, a zero coupon bond. The pricing formula (8.72) becomes

$$99 = 100 e^{-0.5 r(0,0.5)},$$

and we find that

$$r(0,0.5) = 0.0201007. \tag{8.73}$$

For the one year bond, we use formula (8.72) and obtain

$$102 = 2\, e^{-0.5 r(0,0.5)} + 102\, e^{-r(0,1)} = 1.98 + 102\, e^{-r(0,1)}.$$

Solving for $r(0,1)$, we find that

$$r(0,1) \approx 0.0196026. \tag{8.74}$$

For the two year bond, we use the pricing formula (8.72) and obtain

$$103.5 = 2\, e^{-0.5 r(0,0.5)} + 2\, e^{-r(0,1)} + 2\, e^{-1.5 r(0,1.5)} + 102\, e^{-2r(0,2)}. \tag{8.75}$$

While $r(0,0.5)$ and $r(0,1)$ are known from (8.73) and (8.74), two interest rates, $r(0,1.5)$ and $r(0,2)$, are not known. To obtain a unique zero rate curve, we assume that the zero rate $r(0,t)$ is linear for $1 \le t \le 2$, i.e.,

$$r(0,t) = (t-1)\, r(0,2) + (2-t)\, r(0,1), \quad \forall\, 1 \le t \le 2.$$

Then,

$$r(0,1.5) = \frac{r(0,1) + r(0,2)}{2}, \tag{8.76}$$

and (8.75) can be written only in terms of the unknown $x = r(0,2)$ as

$$103.5 = 2\, e^{-0.5 r(0,0.5)} + 2\, e^{-r(0,1)} + 2\, e^{-1.5(r(0,1)+x)/2} + 102\, e^{-2x}. \tag{8.77}$$

This is a nonlinear equation for x which can be solved by Newton's method. After three Newton's method iterations (with initial guess $x_0 = 0.05$, i.e., 5% interest rate), we find that

$$r(0,2) = 0.0219443.$$

Using (8.76), we obtain that $r(0,1.5) = 0.0207735$.

For the five year bond, the pricing formula (8.72) becomes

$$109 = 2 \sum_{i=1}^{9} e^{-0.5i\; r(0,0.5i)} + 102\, e^{-2r(0,5)}. \tag{8.78}$$

While $r(0,0.5)$, $r(0,1)$, $r(0,1.5)$ and $r(0,2)$ are already known, the rest of the zero rates are not. We assume that $r(0,t)$ is linear for $2 \leq t \leq 5$, i.e.,

$$r(0,t) \;=\; \frac{t-2}{3} r(0,5) \;+\; \frac{5-t}{3} r(0,2), \quad \forall\ 2 \leq t \leq 5.$$

Let $x = r(0,5)$. Then,

$$r(0,2.5) \;=\; \frac{x+5r(0,2)}{6}; \; r(0,3) \;=\; \frac{x+2r(0,2)}{3}; \; r(0,3.5) \;=\; \frac{x+r(0,2)}{2};$$

$$r(0,4) \;=\; \frac{2x+r(0,2)}{3}; \; r(0,4.5) \;=\; \frac{5x+r(0,2)}{6}.$$

Formula (8.78) becomes a nonlinear equation in x, which is solved using Newton's method. After four iterations (with initial guess $x_0 = 0.05$), we find that

$$r(0,5) \;=\; 0.0208019.$$

Thus, we found the following zero rates corresponding to the maturities of the four given bonds, i.e., 6 months, 1 year, 2 years, and 5 years:

$$\begin{aligned} r(0,0.5) &= 0.0201007 = 2.0101\%; \\ r(0,1) &= 0.0196026 = 1.9603\%; \\ r(0,2) &= 0.0219443 = 2.1944\%; \\ r(0,5) &= 0.0208019 = 2.0802\%. \end{aligned}$$

Since we assumed that the zero rate curve is linear between any two consecutive bond maturities, we can find $r(0,t)$ for any $0.5 \leq t \leq 5$. Moreover, if we assume that $r(0,t)$ is linear between $t=0$ and $t=0.5$ and if the overnight institutional lending rate $r(0,0)$ is given, then the zero rate $r(0,t)$ is known for any time smaller than the largest maturity of any given bond:

$$r(0,t) \;=\; \begin{cases} 2t\, r(0,0.5) \;+\; (1-2t)\, r(0,0), & \text{if } 0 \leq t \leq 0.5; \\ (2t-1)\, r(0,1) \;+\; 2(1-t)\, r(0,0.5), & \text{if } 0.5 \leq t \leq 1; \\ (t-1)\, r(0,2) \;+\; (2-t)\, r(0,1), & \text{if } 1 \leq t \leq 2; \\ \frac{t-2}{3}\, r(0,5) \;+\; \frac{5-t}{3}\, r(0,2), & \text{if } 2 \leq t \leq 5. \quad \Box \end{cases}$$

8.8 References

While Lagrange multipliers are presented in many calculus books, their treatment usually avoids classifying the critical points of the Lagrangian rigorously. Several examples of applications of Lagrange multipliers can be found

in Protter and Morrey [20]. A brief review of the theory of Lagrange multipliers can be found in Spring [26]. A discussion of the theoretical foundation of Lagrange multipliers for functions of two variables can be found in Edwards [10], and, for functions of n variables, in Stoer and Witzgall [29]. A general linear algebra setup for finding optimal portfolios and its solution using Lagrange multipliers can be found in Ruppert [24].

Optimization problems derived from financial applications and their solutions, including by using Lagrange multipliers, are discussed in great detail in Cornuejols and Tutuncu [7].

More details on numerical methods for solving nonlinear equations can be found in Stoer and Bulirsch [28]. A simple bootstrapping example is given in Hull [14].

8.9 Exercises

1. Find the maximum and minimum of the function $f(x_1, x_2, x_3) = 4x_2 - 2x_3$ subject to the constraints $2x_1 - x_2 - x_3 = 0$ and $x_1^2 + x_2^2 = 13$.

2. Assume that you can trade four assets (and that it is also possible to short the assets). The expected values, standard deviations, and correlations of the rates of return of the assets are:

$$\begin{array}{lll} \mu_1 = 0.08; & \sigma_1 = 0.25; & \rho_{1,2} = -0.25; \\ \mu_2 = 0.12; & \sigma_2 = 0.25; & \rho_{2,3} = -0.25; \\ \mu_3 = 0.16; & \sigma_3 = 0.30; & \rho_{1,3} = 0.25; \\ \mu_4 = 0.05; & \sigma_4 = 0.20; & \rho_{i,4} = 0, \ \ \forall\, i = 1:3. \end{array}$$

(i) Find the asset allocation for a minimal variance portfolio with 12% expected rate of return;

(ii) Find the asset allocation for a maximum expected return portfolio with standard deviation of the rate of return equal to 24%.

3. Use Newton's method to find the yield of a five year semiannual coupon bond with 3.375% coupon rate and price $100\,\frac{1}{32}$. What are the duration and convexity of the bond?

4. Recall that finding the implied volatility from the given price of a call option is equivalent to solving the nonlinear problem $f(x) = 0$, where

$$f(x) = Se^{-qT}N(d_1(x)) - Ke^{-rT}N(d_2(x)) - C$$

and $d_1(x) = \dfrac{\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{x^2}{2}\right)T}{x\sqrt{T}}$, $d_2(x) = \dfrac{\ln\left(\frac{S}{K}\right) + \left(r - q - \frac{x^2}{2}\right)T}{x\sqrt{T}}$.

(i) Show that $\lim_{x\to\infty} d_1(x) = \infty$ and $\lim_{x\to\infty} d_2(x) = -\infty$, and conclude that

$$\lim_{x\to\infty} f(x) = Se^{-qT} - C.$$

(ii) Show that

$$\lim_{x\searrow 0} d_1(x) = \lim_{x\searrow 0} d_2(x) = \begin{cases} -\infty, & \text{if } Se^{(r-q)T} < K; \\ 0, & \text{if } Se^{(r-q)T} = K; \\ \infty, & \text{if } Se^{(r-q)T} > K. \end{cases}$$

(Recall that $F = Se^{(r-q)T}$ is the forward price.)

Conclude that

$$\lim_{x\searrow 0} f(x) = \begin{cases} -C, & \text{if } Se^{(r-q)T} \le K; \\ Se^{-qT} - Ke^{-rT} - C, & \text{if } Se^{(r-q)T} > K \end{cases}$$

(iii) Show that $f(x)$ is a strictly increasing function and

$$\begin{array}{rcccl} -C & < & f(x) & < & Se^{-qT} - C, \ \text{if } Se^{(r-q)T} \le K; \\ Se^{-qT} - Ke^{-rT} - C & < & f(x) & < & Se^{-qT} - C, \ \text{if } Se^{(r-q)T} > K. \end{array}$$

(iv) For what range of call option values does the problem $f(x) = 0$ have a positive solution? Compare your result to the range given in (3.92).

5. A three months at–the–money call on an underlying asset with spot price 30 paying dividends continuously at a 2% rate is worth \$2.5. Assume that the risk free interest rate is constant at 6%.

(i) Compute the implied volatility with six decimal digits accuracy, using the bisection method on the interval [0.0001, 1], the secant method with initial guess 0.5, and Newton's method with initial guess 0.5.

(ii) Let σ_{imp} be the implied volatility previously computed using Newton's method. Use formula (5.77) to compute an approximate value $\sigma_{imp,approx}$ for the implied volatility, and compute the relative error

$$\frac{|\sigma_{imp,approx} - \sigma_{imp}|}{\sigma_{imp}}.$$

6. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$F(x) = \begin{pmatrix} x_1^3 + 2x_1x_2 + x_3^2 - x_2x_3 + 9 \\ 2x_1^2 + 2x_1x_2^2 + x_2^3x_3^2 - x_2^2x_3 - 2 \\ x_1x_2x_3 + x_1^3 - x_3^2 - x_1x_2^2 - 4 \end{pmatrix}.$$

The approximate gradient $\Delta_c F(x) = (\Delta_{c,j} F_i(x))_{i,j=1:n}$ of $F(x)$ is computed using central difference approximations, i.e.,

$$\Delta_{c,j} F_i(x) = \frac{F_i(x + he_j) - F_i(x - he_j)}{2h}, \quad j = 1:n,$$

where e_j is a vector with all entries equal to 0 with the exception of the j-th entry, which is equal to 1.

(i) Solve $F(x) = 0$ using the approximate Newton's algorithm from Table 8.4, by substituting $\Delta_c F(x_{old})$ for $\Delta F(x_{old})$. Use $h = 10^{-6}$, tol_consec $= 10^{-6}$, and tol_approx $= 10^{-9}$, and two different initial guesses:

$$x_0 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \text{and} \quad x_0 = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.$$

(ii) Compare these results to those obtained in section 8.3.2 corresponding to the approximate Newton's method with forward finite difference approximations for $\Delta F(x)$.

7. (i) Use bootstrapping to obtain a zero rate curve from the following prices of Treasury instruments with semiannual coupon payments:

	Coupon Rate	Price
3 – Month T-bill	0	98.7
6 – Month T-bill	0	97.5
2 – Year T-bond	4.875	$100\frac{5}{32}$
3 – Year T-bond	4.875	$100\frac{5}{32}$
5 – Year T-bond	4.625	$99\frac{22}{32}$
10 – Year T-bond	4.875	$101\frac{4}{32}$

Assume that interest is continuously compounded.

(ii) How would the zero rate curves obtained by bootstrapping from the bond prices above, one corresponding to semi-annually compounded interest, and the other one corresponding to continuously computed interest, compare? In other words, will one of the two curves be higher or lower than the other one, and why?

8. Use bootstrapping to obtain a zero rate curve given the prices of the following semiannual coupon bonds:

Maturity	Coupon Rate	Price
6 months	0	97.5
1 year	5	100
20 months	6	103
40 months	5	102
5 years	4	103

Assume that the overnight rate is 5% and that the zero rate curve is linear on the following time intervals:

$$[0, 0.5]; \quad [0.5, 1]; \quad \left[1, \frac{5}{3}\right]; \quad \left[\frac{5}{3}, \frac{10}{3}\right]; \quad \left[\frac{10}{3}, 5\right].$$

Hint: Obtain the zero rate curve $r(0,t)$ for $0 \le t \le 1$ from the prices of the first two bonds, and using the fact that $r(0,0) = 0.05$. The third bond pays coupons in 2, 8, 14, and 20 months, when it also pays the face value of the bond. From the pricing formula (8.72), it follows that

$$103 = 3\, e^{-\frac{1}{6}r(0,\frac{1}{6})} + 3\, e^{-\frac{2}{3}r(0,\frac{2}{3})} + 3\, e^{-\frac{7}{6}r(0,\frac{7}{6})} + 103\, e^{-\frac{5}{3}r(0,\frac{5}{3})}. \quad (8.79)$$

The zero rates $r(0,\frac{1}{6})$ and $r(0,\frac{2}{3})$ are already known. Since the zero rate curve is assumed to be linear for $t \in [1, \frac{5}{3}]$, we find that

$$r(0,t) = \frac{3}{2}\left(\frac{5}{3} - t\right) r(0,1) + \frac{3}{2}(t-1)\, r\left(0, \frac{5}{3}\right).$$

Therefore,

$$r\left(0, \frac{7}{6}\right) = \frac{r(0,1)}{3} + \frac{r\left(0,\frac{5}{3}\right)}{4}.$$

Set $x = r(0,\frac{5}{3})$ and use Newton's method to solve for x in (8.79).

Bibliography

[1] Milton Abramowitz and Irene Stegun. Handbook of Mathematical Functions. National Bureau of Standards, Gaithersburg, Maryland, 10th corrected printing edition, 1970.

[2] Yves Achdou and Olivier Pironneau. Computational Methods for Option Pricing. Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[3] Menachem Brenner and Marti Subrahmanyam. A simple approach to option valuation and hedging in the Black-Scholes model. Financial Analysts Journal, 50(2):25–28, 1994.

[4] Marek Capinski and Ekkehard Kopp. Measure, Integral and Probability. Springer-Verlag London Limited, London, 2nd edition, 2004.

[5] Neil A. Chriss. Black–Scholes and Beyond. Option Pricing Models. McGraw-Hill, Boston, 1997.

[6] Les Clewlow and Chris Strickland. Implementing Derivatives Models. John Wiley & Sons Ltd, Chichester, West Sussex, 1998.

[7] Gerard Cornuejols and Reha Tutuncu. Optimization Methods in Finance. Cambridge University Press, New York, 2007.

[8] Daniel J. Duffy. Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach. John Wiley & Sons Ltd, Chichester, West Sussex, 2006.

[9] Richard Durrett. Probability: Theory and Examples. Duxbury Press, Pacific Grove, California, 3rd edition, 2004.

[10] Henry Edwards. Advanced Calculus of Several Variables. Dover Publications, Mineola, New York, 1995.

[11] Paul Glasserman. Monte Carlo Methods in Financial Engineering. Springer-Verlag New York, Inc., New York, 2004.

[12] Espen G. Haug. The Complete Guide to Option Pricing Formulas. Springer-Verlag New York, Inc., New York, 2004.

[13] John C. Hull. Fundamentals of Futures and Options Markets. Prentice Hall, Upper Saddle River, New Jersey, 5th edition, 2005.

[14] John C. Hull. Options, Futures, and Other Derivatives. Prentice Hall, Upper Saddle River, New Jersey, 6th edition, 2006.

[15] Arieh Iserles. A First Course in the Numerical Analysis of Differential Equations. Cambridge University Press, Cambridge, 1996.

[16] Mark S. Joshi. The Concepts and Practice of Mathematical Finance. Cambridge University Press, New York, 2004.

[17] Harry M. Markowitz. Portfolio selection. Journal of Finance, 1(7):77–91, 1952.

[18] Salih N. Neftci. An Introduction to the Mathematics of the Financial Derivatives. Academic Press, San Diego, California, 2nd edition, 2000.

[19] Salih N. Neftci. Principles of Financial Engineering. Elsevier Academic Press, San Diego, California, 2004.

[20] Murray Protter and Charles Morrey. Intermediate Calculus. Undergraduate Texts in Mathematics. Springer, New York, 2nd edition, 1986.

[21] Sheldon Ross. A First Course in Probability. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 6th edition, 2002.

[22] Halsey Royden. Real Analysis. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 3rd edition, 1988.

[23] Walter Rudin. Real and Complex Analysis. Higher Mathematics Series. McGraw-Hill, Boston, 3rd edition, 1986.

[24] David Ruppert. Statistics and Finance: An Introduction. Undergraduate Studies in Mathematics. Springer Science, New York, 2004.

[25] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 3(19):425–442, 1964.

[26] D. Spring. On the second derivative test for constrained local extrema. American Mathematical Monthly, 92(9):31–643, 1985.

[27] Dan Stefanica. Numerical Linear Algebra Methods for Financial Engineering Applications. FE Press, New York, 2008. To appear.

[28] Josef Stoer and Roland Bulirsch. Introduction to Numerical Analysis. Springer-Verlag New York, Inc., New York, 3rd edition, 2002.

[29] Josef Stoer and Christoph Witzgall. Convexity and Optimization in Finite Dimensions I. Springer-Verlag, Berlin, 1970.

[30] John Strikwerda. Finite Difference Schemes and Partial Differential Equations. Society for Industrial and Applied Mathematics, Philadelphia, 2nd edition, 2004.

[31] Nassim Taleb. Dynamic Hedging: Managing Vanilla and Exotic Options. John Wiley & Sons Ltd, New York, 1997.

[32] Domingo A. Tavella. Quantitative Methods in Derivatives Pricing: An Introduction to Computational Finance. John Wiley & Sons, Inc., New York, 2002.

[33] Paul Wilmott. Paul Wilmott on Quantitative Finance. John Wiley & Sons Ltd, Chichester, West Sussex, 2nd edition, 2006.

[34] Paul Wilmott. Paul Wilmott Introduces Quantitative Finance. John Wiley & Sons Ltd, Chichester, West Sussex, 2007.

[35] Paul Wilmott, Sam Howison, and Jeff Dewynne. The Mathematics of Financial Derivatives: A Student Introduction. Cambridge University Press, New York, 1995.

Index